

Free energy in the generalized Sherrington-Kirkpatrick mean field model.

Dmitry Panchenko*

*Department of Mathematics
Massachusetts Institute of Technology*

Abstract

In [11] Michel Talagrand gave a rigorous proof of the Parisi formula in the classical Sherrington-Kirkpatrick (SK) model. In this paper we build upon the methodology developed in [11] and extend Talagrand's result to the class of SK type models in which the spins have arbitrary prior distribution on a bounded subset of the real line.

Key words: spin glasses.

1 Introduction and main results.

In [11] Michel Talagrand invented a rigorous proof of the Parisi formula for the free energy in the Sherrington-Kirkpatrick model [7]. The methodology developed by Talagrand was based upon a deep extension of Guerra's interpolation method in [2] to coupled systems of spins which provided necessary control of the remainder terms in Guerra's interpolation. The same methodology was successfully used in [13] to compute the free energy in the spherical model and in the present paper we will utilize it in the setting of a generalized Sherrington-Kirkpatrick model in which the prior distribution of the spins is given by an arbitrary probability measure with bounded support on the real line.

Let us start by introducing all necessary notations and definitions. Consider a *bounded* set $\Sigma \subseteq \mathbb{R}$ and a probability measure ν on the Borel σ -algebra on Σ . Given $N \geq 1$, consider a product space (Σ^N, ν^N) which will be called the space of configurations. A configuration $\sigma \in \Sigma^N$ is a vector $(\sigma_1, \dots, \sigma_N)$ of spins σ_i that take values in Σ . For simplicity of notations we will omit index N in ν^N since it will always be clear from the context whether we consider measure ν on Σ or the product measure on Σ^N . For each N we consider a Hamiltonian $H_N(\sigma)$ on Σ^N that is a Gaussian process indexed by $\sigma \in \Sigma^N$. We will assume that $H_N(\sigma)$ is jointly measurable in (σ, g) , where g is the generic point of the underlying probability space

*Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139 email: panchenk@math.mit.edu

on which the Gaussian process $H_N(\boldsymbol{\sigma})$ is defined. We assume that for a certain sequence $c(N) \rightarrow 0$ and a certain function $\xi : \mathbb{R} \rightarrow \mathbb{R}$ we have

$$\forall \boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2 \in \Sigma^N, \quad \left| \frac{1}{N} \mathbb{E} H_N(\boldsymbol{\sigma}^1) H_N(\boldsymbol{\sigma}^2) - \xi(R_{1,2}) \right| \leq c(N), \quad (1.1)$$

where

$$R_{1,2} = \frac{1}{N} \sum_{i \leq N} \sigma_i^1 \sigma_i^2 \quad (1.2)$$

is called *the overlap* of the configurations $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2$. We will assume that $\xi(x)$ is three times continuously differentiable and satisfies the following conditions

$$\xi(0) = 0, \quad \xi(x) = \xi(-x), \quad \xi''(x) > 0 \text{ if } x > 0. \quad (1.3)$$

We will denote *the self-overlap* of $\boldsymbol{\sigma}$ by

$$R_{1,1} = \frac{1}{N} \sum_{i \leq N} (\sigma_i)^2.$$

One defines the Gibbs measure G_N on Σ^N by

$$dG_N(\boldsymbol{\sigma}) = \frac{1}{Z_N} \exp H_N(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}) \quad (1.4)$$

where the normalizing factor

$$Z_N = \int_{\Sigma^N} \exp H_N(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma})$$

is called the *partition function*. This definition of the Gibbs measure also includes the case of models with external field because given a measurable function $h(\sigma)$ on Σ and the Gibbs measure defined by

$$dG_N(\boldsymbol{\sigma}) = \frac{1}{Z_N} \exp \left(H_N(\boldsymbol{\sigma}) + \sum_{i \leq N} h(\sigma_i) \right) d\nu(\boldsymbol{\sigma})$$

we can simply make the change of measure $d\nu'(\sigma) \sim \exp h(\sigma) d\nu(\sigma)$ to represent this Gibbs measure as (1.4). The only assumption that we need to make on $h(\sigma)$ is that

$$\int \exp h(\sigma) d\nu(\sigma) < \infty$$

which holds, for example, when $h(\sigma)$ is uniformly bounded. This general model includes the original Sherrington-Kirkpatrick (SK) model in [7] and the Ghatak-Sherrington (GS) model in [1] that will be considered in more detail in Section 2.1. In both cases the Hamiltonian $H_N(\boldsymbol{\sigma})$ is given by

$$H_N(\boldsymbol{\sigma}) = \frac{\beta}{\sqrt{N}} \sum_{i < j} g_{ij} \sigma_i \sigma_j \quad (1.5)$$

for some $\beta > 0$ and i.i.d. Gaussian r.v. g_{ij} , the set Σ is equal to $\{-1, +1\}$ in the SK model and $\{0, \pm 1, \dots, \pm S\}$ for some integer S in the GS model and in both cases the measure ν is uniform on Σ . In the case of the SK model the function $h(\sigma)$ is given by $h(\sigma) = h\sigma$ with the *external field* parameter $h \in \mathbb{R}$, and in the case of the GS model it is given by $h(\sigma) = h\sigma^2$ with the *crystal field* parameter $h \in \mathbb{R}$. Let us define

$$F_N = \frac{1}{N} \mathbb{E} \log Z_N = \frac{1}{N} \mathbb{E} \log \int_{\Sigma^N} \exp H_N(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}), \quad (1.6)$$

which (usually, with the factor $-\beta^{-1}$ which we omit for simplicity of notations) is called the *free energy* of the system (Σ^N, G_N) . The main goal of this paper is to find the limit $\lim_{N \rightarrow \infty} F_N$. It will soon become clear that the main difference of the above model from the classical SK model lies in the fact that in the classical model the length of any configuration $\boldsymbol{\sigma} \in \{-1, +1\}^N$ was constant, $|\boldsymbol{\sigma}| = \sqrt{N}$, which is not always true here. In general, if Σ is not of the type $\{-a, +a\}$ for some $a \in \mathbb{R}$ then the length of the configuration or self-overlap $R_{1,1} = |\boldsymbol{\sigma}|^2/N$ will become variable. As a result, in order to make the methodology of Guerra and Talagrand work, we will first have to compute the *local* free energy of the set of configurations with constrained self-overlap.

Let us now describe the analogue of the Parisi formula that gives the limit of (1.6). Let $[d, D]$ be the smallest interval such that

$$\nu(\{\sigma : \sigma^2 \in [d, D]\}) = 1. \quad (1.7)$$

In other words, $d \leq \sigma^2 \leq D$ with probability one and σ^2 can take values arbitrarily close to d and D with positive probability. From now on we will simply say that $d \leq \sigma^2 \leq D$ for all $\sigma \in \Sigma$. Let us consider $u \in [d, D]$ and a sequence (ε_N) such that $\varepsilon_N > 0$ and $\lim_{N \rightarrow \infty} \varepsilon_N = 0$, and consider a sequence of sets

$$U_N = \left\{ \boldsymbol{\sigma} \in \Sigma^N : R_{1,1} \in [u - \varepsilon_N, u + \varepsilon_N] \right\}. \quad (1.8)$$

We define

$$F_N(u, \varepsilon_N) = \frac{1}{N} \mathbb{E} \log Z_N(u, \varepsilon_N), \quad (1.9)$$

where

$$Z_N(u, \varepsilon_N) = \int_{U_N} \exp H_N(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}).$$

$F_N(u, \varepsilon_N)$ is the free energy of the subset of configurations in (1.8). We will first compute $\lim_{N \rightarrow \infty} F_N(u, \varepsilon_N)$ for some sequence (ε_N) for each $u \in [d, D]$. Consider an integer $k \geq 1$, numbers

$$0 = m_0 \leq m_1 \leq \dots \leq m_{k-1} \leq m_k = 1 \quad (1.10)$$

and, given $u \in [d, D]$,

$$0 = q_0 \leq q_1 \leq \dots \leq q_k \leq q_{k+1} = u. \quad (1.11)$$

We will write $\mathbf{m} = (m_0, \dots, m_k)$ and $\mathbf{q} = (q_0, \dots, q_{k+1})$. Consider independent centered Gaussian r.v. z_p for $0 \leq p \leq k$ with

$$\mathbb{E} z_p^2 = \xi'(q_{p+1}) - \xi'(q_p). \quad (1.12)$$

Given $\lambda \in \mathbb{R}$, we define the r.v.

$$X_{k+1} = \log \int_{\Sigma} \exp \left(\sigma \sum_{0 \leq p \leq k} z_p + \lambda \sigma^2 \right) d\nu(\sigma) \quad (1.13)$$

and, recursively for $l \geq 0$, define

$$X_l = \frac{1}{m_l} \log \mathbb{E}_l \exp m_l X_{l+1}, \quad (1.14)$$

where \mathbb{E}_l denotes the expectation in the r.v. $(z_p)_{p \geq l}$. When $m_l = 0$ this means $X_l = \mathbb{E}_l X_{l+1}$. Clearly, $X_0 = X_0(\mathbf{m}, \mathbf{q}, \lambda)$ is a non-random function of the parameters \mathbf{m}, \mathbf{q} and λ . Whenever it does not create ambiguity we will keep this dependence implicit. Let us note that X_0 also depends on u through \mathbf{q} since in (1.11) we have $q_{k+1} = u$. Let

$$\mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u) = -\lambda u + X_0(\mathbf{m}, \mathbf{q}, \lambda) - \frac{1}{2} \sum_{1 \leq l \leq k} m_l (\theta(q_{l+1}) - \theta(q_l)) \quad (1.15)$$

where $\theta(q) = q\xi'(q) - \xi(q)$, and define

$$\mathcal{P}(\xi, u) = \inf \mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u), \quad (1.16)$$

where the infimum is taken over all λ, k, \mathbf{m} and \mathbf{q} . Finally, we define

$$\mathcal{P}(\xi) = \sup_{d \leq u \leq D} \mathcal{P}(\xi, u). \quad (1.17)$$

We will first prove the following.

Theorem 1 *Given $u \in [d, D]$ and a sequence $(\varepsilon_N)_{N \geq 1}$ that goes to zero slowly enough,*

$$\lim_{N \rightarrow \infty} F_N(u, \varepsilon_N) = \mathcal{P}(\xi, u). \quad (1.18)$$

Gaussian concentration of measure will imply that the limit of the global free energy can be computed by maximizing the local free energy.

Theorem 2 *We have*

$$\lim_{N \rightarrow \infty} F_N = \mathcal{P}(\xi). \quad (1.19)$$

Organization of the paper. In Section 2 we describe the replica symmetric region of the model and discuss the example of the Ghatak-Sherrington model. In Section 3 we introduce the construction (which we call the Parisi functional) that is used often throughout the paper and study some of its properties. In Section 4 we prove the analogue of Guerra's interpolation and explain why it seems to be necessary to impose the constraint on the self-overlap in order to utilize the methodology of Talagrand in [11]. Compared to the classical SK model where this problem does not occur, for the general model considered in this paper a brand new argument is required to remove the constraint on the self-overlap at the end of Guerra's interpolation. This constitutes a certain nontrivial large deviation problem that is

solved in Section 5. In Section 6 we show how Theorem 1 can be reduced to certain apriori estimates on the error terms in Guerra's interpolation. For the most part, the proof of these apriori estimates goes along the lines of the methodology developed by Talagrand in [11] but, nonetheless, considerable effort is required to verify that the arguments and numerous computations in [11] extend to this more general model. We carry out these computations in Appendix A. In Section 7 we show how the global Parisi formula of Theorem 2 follows from the local Parisi formula of Theorem 1 and a certain concentration of measure result. Finally, certain values of the parameter u in Theorem 1 require small modifications of some arguments but, fortunately, these cases can be reduced to the classical model considered in [11], a work which is postponed until Appendix B.

2 Replica symmetric region.

In this section we will describe a relatively simple necessary and sufficient condition in terms of the parameters of the model which guarantees that the infimum on the right hand side of (1.16) is achieved when $k = 1$. If this happens then $\mathcal{P}(\xi, u)$ will be called a local replica symmetric solution. In Section B we will explain that the cases when $u = d$ or $u = D$ in Theorem 1 can be reduced to the classical SK model for which the domain of validity of the replica symmetric solution was described in [8] and, hence, in the rest of the paper we will assume that

$$d < u < D. \quad (2.1)$$

If the infimum in (1.16) is achieved when $k = 1$ then

$$\mathbf{m} = (0, 1), \quad \mathbf{q} = (0, q, u) \text{ for some } q \in [0, u] \quad (2.2)$$

and λ and q are the only variables in $\mathcal{P}_1(\mathbf{m}, \mathbf{q}, \lambda, u)$ which, hence, can be written as

$$\mathcal{P}_1(q, \lambda) = -\lambda u - \frac{1}{2}(\theta(u) - \theta(q)) + \mathbb{E} \log \int_{\Sigma} \exp H(\sigma) d\nu(\sigma), \quad (2.3)$$

where

$$H(\sigma) = \sigma z_0 + \lambda \sigma^2 + \frac{1}{2} \sigma^2 (\xi'(u) - \xi'(q)).$$

Let us define the (local) replica symmetric solution by

$$\text{RS}(u) = \inf_{\lambda, q} \mathcal{P}_1(q, \lambda) \quad (2.4)$$

and describe the criterion which guarantees that $\mathcal{P}(\xi, u) = \text{RS}(u)$. We will prove that if (2.1) holds then the infimum on the right hand side of (2.4) is achieved on some λ and q which, therefore, must satisfy the critical point conditions

$$\frac{\partial \mathcal{P}_1}{\partial \lambda} = \frac{\partial \mathcal{P}_1}{\partial q} = 0. \quad (2.5)$$

From now on let (q, λ) be such a pair, i.e. $\text{RS}(u) = \mathcal{P}_1(q, \lambda)$. Suppose that $\mathcal{P}(\xi, u) = \text{RS}(u)$. Then taking $k = 2$ in (1.16) should not decrease the infimum on the right hand side. Let us take

$$k = 2, \quad \mathbf{m} = (0, m, 1) \quad \text{and} \quad \mathbf{q} = (0, q, a, u) \quad \text{for} \quad a \in [q, u]. \quad (2.6)$$

With this choice of parameters $\mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u)$ becomes

$$\Phi(m, a) = -\lambda u - \frac{1}{2}m(\theta(a) - \theta(q)) - \frac{1}{2}(\theta(u) - \theta(a)) + \frac{1}{m}\mathbb{E} \log \mathbb{E}_1 X^m, \quad (2.7)$$

where

$$X = \int_{\Sigma} \exp H'(\sigma) d\nu(\sigma) \quad \text{and} \quad H'(\sigma) = \sigma(z_0 + z_1) + \lambda \sigma^2 + \frac{1}{2} \sigma^2 (\xi'(u) - \xi'(a)) \quad (2.8)$$

and where $\mathbb{E} z_0^2 = \xi'(q)$ and $\mathbb{E} z_1^2 = \xi'(a) - \xi'(q)$. It should be obvious that $\Phi(1, a) = \mathcal{P}_1(q, \lambda)$ for any $q \leq a \leq u$. The derivative of $\Phi(m, a)$ with respect to m at $m = 1$ can not be positive because, otherwise, by decreasing m slightly we could decrease $\Phi(m, a)$ that would imply

$$\mathcal{P}(\xi, u) \leq \Phi(m, a) < \mathcal{P}_1(q, \lambda) = \text{RS}(u).$$

A simple computation gives

$$f(a) = \frac{\partial \Phi}{\partial m}(m, a) \Big|_{m=1} = -\frac{1}{2}(\theta(a) - \theta(q)) + \mathbb{E} \frac{X}{\mathbb{E}_1 X} \log \frac{X}{\mathbb{E}_1 X} \quad (2.9)$$

and, hence, the following condition is necessary if $\mathcal{P}(\xi, u) = \text{RS}(u)$,

$$f(a) \leq 0 \quad \text{for all} \quad q \leq a \leq u. \quad (2.10)$$

This derivative $f(a)$ represents what is usually called the replica symmetry breaking fluctuations. Let us note that since $\Phi(m, q) = \mathcal{P}_1(q, \lambda)$ does not depend on m , we have $f(q) = 0$. Also, it is easy to check that (2.5) implies that $f'(q) = 0$. Therefore, if (2.10) holds then we must have

$$f''(q) \leq 0, \quad (2.11)$$

which in the SK model is called the *Almeida-Thouless* condition. It is believed (and numerical computations show) that in the classical SK model (2.11) implies (2.10). However, we will give an example below where this is not the case and, therefore, condition (2.10) can not be weakened to (2.11) in general. We will prove that (2.10) is (necessary and) sufficient for $\mathcal{P}(\xi, u) = \text{RS}(u)$.

Theorem 3 *If a pair (λ, q) satisfies (2.5) and (2.10) then $\mathcal{P}(\xi, u) = \mathcal{P}_1(q, \lambda)$ and such pair (λ, q) is unique.*

The proof of this theorem goes in parallel with the proof of Theorem 1 as will be explained in Section 6. However, its proof would be immediate if we knew that the functional \mathcal{P}_k defined in (1.15) was convex in \mathbf{m} . The conjecture that \mathcal{P}_k is indeed convex in \mathbf{m} was made in [14] and in [6] where a partial result was proved. We do not give the details here but, shortly speaking, the convexity of \mathcal{P}_k would imply the uniqueness of the minimum in the optimization problem (1.16) and since (2.10) means that the replica symmetric choice of parameters is a local minimum in (1.16), hence, it would be a global minimum.

2.1 Ghatak-Sherrington model.

Let us consider the Ghatak-Sherrington model introduced in [1] with $\Sigma = \{-1, 0, +1\}$, the Hamiltonian $H_N(\sigma)$ defined in (1.5), the measure ν is the counting measure on Σ and the external field $h(\sigma) = h\sigma^2$ for some $h \in \mathbb{R}$. This choice of parameters gives

$$\mathcal{P}_1(q, \lambda) = -\lambda u - \frac{\beta^2}{4}(u^2 - q^2) + \mathbb{E} \log \left(1 + 2\text{ch}(z\beta\sqrt{q}) \exp\left(\lambda + h + \frac{1}{2}\beta^2(u - q)\right) \right),$$

where z is a standard normal r.v. and we keep the dependence of \mathcal{P}_1 on u implicit. Because of the symmetry of the model, rather than the replica symmetric solution, one is usually interested in the case when $\mathcal{P}(\xi, u) = \mathcal{P}_1(q, \lambda)$ for $q = 0$ and some λ . It is easy to check that

$$\frac{\partial \mathcal{P}_1}{\partial q} = \frac{\beta^2 q}{2} - \frac{\beta^2}{2} \mathbb{E} \left(\frac{2\text{sh}(z\beta\sqrt{q}) \exp(\lambda + h + \beta^2(u - q)/2)}{1 + 2\text{ch}(z\beta\sqrt{q}) \exp(\lambda + h + \beta^2(u - q)/2)} \right)^2$$

and, therefore, $q = 0$ always satisfies the critical point condition (2.5). If $\mathcal{P}(\xi, u) = \mathcal{P}_1(0, \lambda)$ for some λ then we will call $\mathcal{P}(\xi, u)$ a *paramagnetic solution* and denote it by $\text{PM}(u)$.

$$\begin{aligned} \text{PM}(u) = \inf_{\lambda} \mathcal{P}_1(0, \lambda) &= \inf_{\lambda} \left(-\lambda u - \frac{1}{4}\beta^2 u^2 + \log(1 + 2e^{\lambda+h+\frac{\beta^2}{2}u}) \right) \\ &= hu + \frac{1}{4}\beta^2 u^2 + \inf_{\lambda'} \left(-\lambda' u + \log(1 + 2e^{\lambda'}) \right) \\ &= hu + \frac{1}{4}\beta^2 u^2 + u \log \frac{2}{u} + (1 - u) \log \frac{1}{1 - u}, \end{aligned}$$

where we made the change of variable $\lambda' = \lambda + h + \beta^2 u/2$. The infimum is achieved on

$$\lambda = -h - \frac{1}{2}\beta^2 u + \log \frac{u}{2(1 - u)}. \quad (2.12)$$

Theorem 3 can be applied to this model to describe when the local free energy $\mathcal{P}(\xi, u)$ is given by the paramagnetic solution $\text{PM}(u)$. When $q = 0$, the definition (2.8) implies that

$$X = 1 + 2e^{\lambda+h+\frac{\beta^2}{2}(u-a)} \text{ch}(z_1\beta\sqrt{a}) = 1 + \frac{u}{1-u} e^{-\frac{\beta^2}{2}a} \text{ch}(z_1\beta\sqrt{a}).$$

Using the fact that

$$\mathbb{E} e^{-\frac{\beta^2}{2}a} \text{ch}(z\beta\sqrt{a}) = 1 \quad (2.13)$$

we get $\mathbb{E}_1 X = 1 + u/(1 - u) = 1/(1 - u)$ and for $0 \leq a \leq u$,

$$f(a) = -\frac{1}{4}\beta^2 a^2 + \mathbb{E} \left(1 - u + u e^{-\frac{\beta^2}{2}a} \text{ch}(z\beta\sqrt{a}) \right) \log \left(1 - u + u e^{-\frac{\beta^2}{2}a} \text{ch}(z\beta\sqrt{a}) \right). \quad (2.14)$$

Since λ in (2.12) and $q = 0$ satisfy (2.5), Theorem 3 implies that the subset of configurations with constrained self-overlap $R_{1,1} \approx u$ will be in the paramagnetic phase, $\mathcal{P}(\xi, u) = \text{PM}(u)$, if and only if

$$f(a) \leq 0 \quad \text{for } 0 \leq a \leq u. \quad (2.15)$$

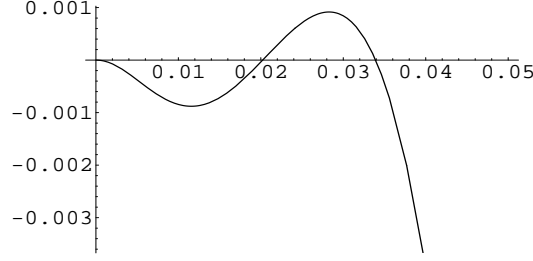


Figure 1: A function $f(a)$ for $u = 0.05, \beta = 17.5$ and $a \leq u$.

It is easy to check that $f''(0) = \beta^2(-1 + \beta^2 u^2)/2$ and, therefore, (2.11) implies

$$-1 + \beta^2 u^2 \leq 0 \quad \text{or} \quad \beta u \leq 1. \quad (2.16)$$

It is tempting to conjecture that (2.16) implies (2.15) but, unfortunately, even though it is expected to be true in the classical SK model it is not always true here. For example, one can check that for $u = 0.05$ and $\beta = 17.5$, (2.16) holds but (2.15) fails (see figure 1).

Next, we would like to describe the region of parameters β and h such that the system as a whole is in the paramagnetic phase in the sense that

$$\mathcal{P}(\xi) = \sup_{u \in [0,1]} \mathcal{P}(\xi, u) = \mathcal{P}(\xi, u') = \text{PM}(u'), \quad (2.17)$$

i.e. the local free energy $\mathcal{P}(\xi, u)$ is maximized at some point u' where $\mathcal{P}(\xi, u') = \text{PM}(u')$ and, thus, the global free energy is given by the paramagnetic solution, $\mathcal{P}(\xi) = \text{PM}(u')$. In figure 2 we show a phase diagram in the coordinates $(h\beta^{-1}, \beta^{-1})$ to compare it with [1] where the phase diagram was given in these coordinates. According to [1], Regions 1 and 3 constitute the paramagnetic phase where (2.17) holds, and Region 2 is the spin glass phase where (2.17) fails, i.e. $\mathcal{P}(\xi, u') < \text{PM}(u')$ for u' such that $\sup_u \mathcal{P}(\xi, u) = \mathcal{P}(\xi, u')$. We will explain how these regions were defined in [1] and argue that in Regions 1 and 3 (2.17) holds. We consider Regions 1 and 3 separately because Region 1 can be treated rigorously.

The way figure 2 was obtained in [1] is apparently as follows. The authors used the replica symmetric approximation which means that instead of looking at the free energy $\mathcal{P}(\xi)$ which by Theorem 2 is given by $\sup_u \mathcal{P}(\xi, u)$ they considered a replica symmetric approximation of $\mathcal{P}(\xi)$ given by

$$\sup_u \text{RS}(u) = \sup_u \inf_{q, \lambda} \mathcal{P}_1(q, \lambda). \quad (2.18)$$

The value provided by this approximation, in general, is not equal to the actual free energy $\mathcal{P}(\xi)$ and is only an upper bound. However, this optimization problem is much easier than the case of the general Parisi formula in Theorem 2 since (2.18) depends only on three parameters (u, q, λ) . The saddle point conditions for the solution (u, q, λ) of (2.18) are given by

$$\frac{\partial \mathcal{P}_1}{\partial \lambda} = 0, \quad \frac{\partial \mathcal{P}_1}{\partial q} = 0 \quad \text{and} \quad \lambda = 0, \quad (2.19)$$

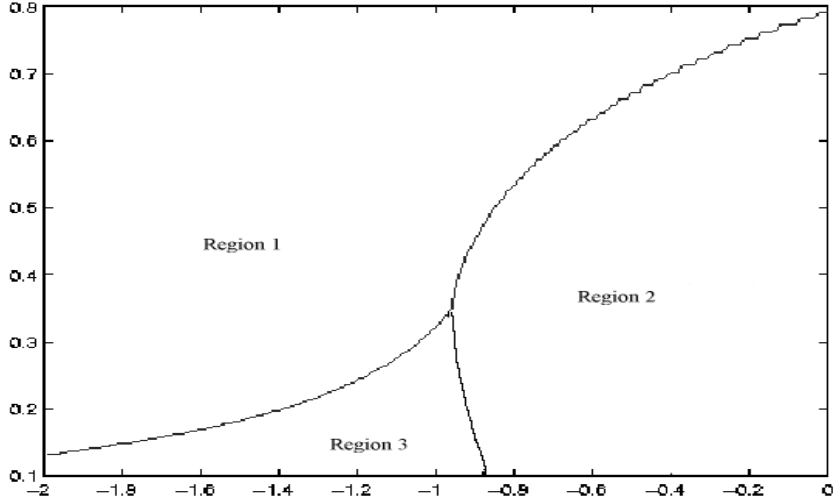


Figure 2: A phase diagram in $(h\beta^{-1}, \beta^{-1})$ coordinates. Regions 1 and 3 are a paramagnetic phase and, according to [1], Region 2 is a spin glass phase.

since it is easy to check that maximizing over u and using $\partial\mathcal{P}_1/\partial\lambda = 0$ gives $\lambda = 0$. Hence, (2.19) reduces to solving the system of two equations. It was predicted in [1] that the paramagnetic phase coincides with the set of parameters (β, h) for which the infimum in (2.18) is attained at the saddle point (u, q, λ) such that $q = 0$. This set is given by the union of Regions 1 and 3. On the complement, Region 2, the replica symmetric approximation of the free energy $\mathcal{P}(\xi)$ is given by $\mathcal{P}_1(q, \lambda)$ with $q \neq 0$ and the authors in [1] concluded that it is, therefore, a spin glass phase in the sense that (2.17) fails. However, this conclusion in general requires further justification because (2.18) is only an approximation of the general Parisi formula.

Region 1. We will define Region 1 below after we explain a simple but important property of this model. The fact that Region 1 is a paramagnetic phase will follow from this property. First, let us observe that for any fixed a the function $f_u(a) = f(a)$ is increasing in u , where for a moment we made the dependence on u explicit. We have

$$\frac{\partial f_u(a)}{\partial u} = \mathbb{E}\left(-1 + e^{-\frac{\beta^2}{2}a} \text{ch}(z\beta\sqrt{a})\right) \log(1 - u + ue^{-\frac{\beta^2}{2}a} \text{ch}(z\beta\sqrt{a})).$$

It is easy to check that for any $u \in [0, 1]$, the function $x \rightarrow (-1 + x) \log(1 - u + ux)$ is convex for $x \geq 0$ and, therefore, (2.13) and Jensen's inequality imply that $\partial f_u(a)/\partial u \geq 0$. This means that if we take $u_1 \leq u_2$ then

$$f_{u_2}(a) \leq 0 \text{ for } a \leq u_2 \implies f_{u_1}(a) \leq 0 \text{ for } a \leq u_1. \quad (2.20)$$

If for some $u_2 \in [0, 1]$ we have $\mathcal{P}(\xi, u_2) = \text{PM}(u_2)$ then (2.15) holds for $u = u_2$. By (2.20), (2.15) holds for any $u = u_1 \leq u_2$ and, therefore, $\mathcal{P}(\xi, u_1) = \text{PM}(u_1)$. This proves the following important property: for any β and h there exists $u_0 = u_0(\beta) \in [0, 1]$ such that

$$\mathcal{P}(\xi, u) = \text{PM}(u) \text{ if and only if } u \leq u_0. \quad (2.21)$$

Suppose that the maximum of $\text{PM}(u)$ is achieved on some $u_1 \leq u_0$. Then the system as a whole will be in the paramagnetic phase because

$$\sup_{u \in [0,1]} \mathcal{P}(\xi, u) \geq \sup_{u \leq u_0} \mathcal{P}(\xi, u) = \sup_{u \leq u_0} \text{PM}(u) = \sup_{u \in [0,1]} \text{PM}(u) \geq \sup_{u \in [0,1]} \mathcal{P}(\xi, u) \quad (2.22)$$

and, therefore,

$$\mathcal{P}(\xi) = \sup_{u \in [0,1]} \mathcal{P}(\xi, u) = \sup_{u \leq u_0} \text{PM}(u) = \text{PM}(u_1). \quad (2.23)$$

Region 1 is precisely where the maximum of $\text{PM}(u)$ is achieved on some $u_1 \leq u_0$ and, thus, it is a subset of the paramagnetic phase.

Region 3. As we mentioned above, in Region 3 the optimization problem (2.18) is solved at the saddle point (u', q', λ') such that $q' = 0$. This means that for this u' ,

$$\sup_u \text{RS}(u) = \text{RS}(u') = \inf_{q, \lambda} \mathcal{P}_1(q, \lambda) = \inf_{\lambda} \mathcal{P}_1(0, \lambda) = \text{PM}(u'). \quad (2.24)$$

By itself this fact does allow us to conclude that we are in the paramagnetic phase, but apparently in this particular model when this happens we also have $u' \leq u_0$ where u_0 was defined in (2.21) and we do not see how to prove this using calculus. Should this numerical observation reflect the true situation, as seems likely, then (2.24) would imply that

$$\text{PM}(u') = \mathcal{P}(\xi, u') \leq \sup_u \mathcal{P}(\xi, u) \leq \sup_u \text{RS}(u) = \text{PM}(u')$$

and we would again be in the paramagnetic phase.

Region 2. The fact that the infimum in (2.18) is attained at the saddle point (u', q', λ') with $q' \neq 0$ implies that $u' > u_0$ and

$$\sup_{u \leq u_0} \text{PM}(u) < \text{RS}(u'). \quad (2.25)$$

However, since $\text{RS}(u)$ is only an upper bound on $\mathcal{P}(\xi, u)$, in general, (2.25) does not exclude the possibility that

$$\mathcal{P}(\xi, u) < \sup_{u \leq u_0} \text{PM}(u) \quad \text{for all } u > u_0$$

which would imply (2.17). At this moment we do not see how to prove that Region 2 is a spin glass phase except by checking directly that (2.17) fails.

3 Parisi functional.

We will often consider iterative constructions similar to (1.13) and (1.14), so it will be convenient to define an operator that implements this recursion. In each case we will only need to specify the parameters of the operator. We will call this operator the Parisi functional. Given $k \geq 1$, consider a vector

$$\mathbf{m} = (m_0, \dots, m_k) \quad (3.1)$$

such that all coordinates $m_i \geq 0$. Let $\mathbf{z} = (z_l)_{0 \leq l \leq k}$ be a collection of independent random vectors.

Suppose that we are given a random variable F that is a function of \mathbf{z} , $F = F(\mathbf{z})$. Then we let $F_{k+1} = F$ and for $0 \leq l \leq k$ define iteratively

$$F_l = \frac{1}{m_l} \log \mathbb{E}_l \exp m_l F_{l+1}, \quad (3.2)$$

where \mathbb{E}_l denotes the expectation in $(z_i)_{i \geq l}$. When $m_l = 0$, this means $F_l = \mathbb{E}_l F_{l+1}$.

Definition. We define the Parisi functional by

$$\mathcal{P}(\mathbf{m})F = F_0. \quad (3.3)$$

With these notations the definition of X_0 given by (1.10) - (1.14) can be written as

$$X_0 = \mathcal{P}(\mathbf{m})X_{k+1}. \quad (3.4)$$

Let us describe several immediate properties of the Parisi functional that will be often used throughout the paper. It is obvious by induction in (3.2) that for any constant $c \in \mathbb{R}$,

$$\mathcal{P}(\mathbf{m})(c + F) = c + \mathcal{P}(\mathbf{m})F. \quad (3.5)$$

Similarly, if $\mathbf{z}^j = (z_p^j)_{0 \leq p \leq k}$ are independent for $j = 1, 2$ and $F^j = F^j(\mathbf{z}^j)$ then

$$\mathcal{P}(\mathbf{m})(F^1 + F^2) = \mathcal{P}(\mathbf{m})F^1 + \mathcal{P}(\mathbf{m})F^2. \quad (3.6)$$

If $F \leq F'$ then

$$\mathcal{P}(\mathbf{m})F \leq \mathcal{P}(\mathbf{m})F'. \quad (3.7)$$

The next property plays an important role in Talagrand's interpolation for two copies of the system. Suppose that we have two random variables

$$F^j = F\left(\sum_{p \leq k} z_p^j\right) \text{ for } j = 1, 2 \quad (3.8)$$

such that

$$z_l^1 = z_l^2 \text{ for } l < r \text{ and } z_l^1 \text{ and } z_l^2 \text{ are independent copies for } l \geq r. \quad (3.9)$$

Let us define $\mathbf{n} = (n_p)_{p \leq k}$ such that

$$n_l = \frac{m_l}{2} \text{ for } l < r \text{ and } n_l = m_l \text{ for } l \geq r. \quad (3.10)$$

Lemma 1 *Given (3.8), (3.9) and (3.10) let $F = F^1 + F^2$ and define F_l^j by (3.2) and F_l by (3.2) with \mathbf{m} replaced by \mathbf{n} . Then,*

$$F_l = F_l^1 + F_l^2 \text{ for } l > r \text{ and } F_l = 2F_l^1 = 2F_l^2 \text{ for } l \leq r. \quad (3.11)$$

In particular,

$$\mathcal{P}(\mathbf{n})(F^1 + F^2) = 2\mathcal{P}(\mathbf{m})F^1. \quad (3.12)$$

Proof. The proof follows by induction in (3.2). For $l > r$, using independence of z_l^1 and z_l^2 we get

$$F_l = \frac{1}{n_l} \log \mathbb{E}_l \exp n_l F_{l+1} = \frac{1}{m_l} \log \mathbb{E}_l \exp m_l (F_{l+1}^1 + F_{l+1}^2) = F_l^1 + F_l^2.$$

For $l = r$ all independent copies already have been averaged and since $z_l^1 = z_l^2$ for $l < r$, $F_r^1 = F_r^2$ and $F_r = 2F_r^1$. Finally, by induction for $l < r$ we get

$$F_l = \frac{1}{n_l} \log \mathbb{E}_l \exp n_l F_{l+1} = \frac{2}{m_l} \log \mathbb{E}_l \exp \frac{m_l}{2} 2F_{l+1}^1 = 2F_l^1$$

which for $l = 0$ proves (3.12). □

The next property concerns the computation of the derivatives of $\mathcal{P}(\mathbf{m})F$. With the notations of (3.1) and (3.2), let us define

$$W_l = \exp m_l (F_{l+1} - F_l). \quad (3.13)$$

Note that by definition of F_l in (3.2) we have $\mathbb{E}_l W_l = 1$. Also, since F_l , F_{l+1} and W_l do not depend on z_i for $i \geq l + 1$, we can write

$$\mathbb{E}_l W_l W_{l+1} = \mathbb{E}_l \mathbb{E}_{l+1} W_l W_{l+1} = \mathbb{E}_l W_l \mathbb{E}_{l+1} W_{l+1} = 1.$$

Repeating the same argument,

$$\mathbb{E}_l W_l \dots W_k = 1. \quad (3.14)$$

This fact will be used often below.

Lemma 2 *For a generic variable x ,*

$$\frac{\partial \mathcal{P}(\mathbf{m})}{\partial x} = \mathbb{E}_0 W_0 \dots W_k \frac{\partial F}{\partial x}. \quad (3.15)$$

Proof. By (3.2), $\exp m_l F_l = \mathbb{E}_l \exp m_l F_{l+1}$ and, therefore,

$$m_l \exp m_l F_l \frac{\partial F_l}{\partial x} = m_l \mathbb{E}_l \exp m_l F_{l+1} \frac{\partial F_{l+1}}{\partial x}.$$

Since F_l does not depend on z_i for $i \geq l$,

$$\frac{\partial F_l}{\partial x} = \mathbb{E}_l W_l \frac{\partial F_{l+1}}{\partial x},$$

where W_l was defined in (3.13). Applying the same equation to F_{l+1} gives

$$\frac{\partial F_l}{\partial x} = \mathbb{E}_l W_l \mathbb{E}_{l+1} W_{l+1} \frac{\partial F_{l+2}}{\partial x} = \mathbb{E}_l W_l W_{l+1} \frac{\partial F_{l+2}}{\partial x}$$

since F_l , F_{l+1} and, therefore, W_l do not depend on (z_i) for $i \geq l + 1$. Repeating the same argument inductively we get

$$\frac{\partial F_l}{\partial x} = \mathbb{E}_l W_l \dots W_k \frac{\partial F}{\partial x}, \quad (3.16)$$

which for $l = 0$ implies (3.15). □

We will often assume that the coordinates of \mathbf{m} are arranged in a nondecreasing order,

$$m_0 \leq \dots \leq m_k. \quad (3.17)$$

The following lemma provides a useful control of the expressions of the type (3.15) that appear as the derivatives of the Parisi functional.

Lemma 3 *Suppose (3.17) holds. Let $f = f(\mathbf{z})$ and $F = F(\mathbf{z})$ and let W_l be defined by (3.13). Suppose that $f \leq F$ and let $m = m_r$ be the first nonzero element in (3.17). Then*

$$\mathbb{E}_0 \log \mathbb{E}_r W_r \dots W_k \exp m_k(f - F) \leq m(\mathcal{P}(\mathbf{m})f - \mathcal{P}(\mathbf{m})F). \quad (3.18)$$

Proof. Let us define $U = \exp m_k(f - F)$. Using (3.13), we can write

$$W_k U = \exp m_k(f - F_k)$$

and since F_k does not depend on z_k , this implies that

$$\mathbb{E}_k W_k U = \exp(-m_k F_k) \mathbb{E}_k \exp m_k f = \exp m_k(f_k - F_k).$$

We will proceed by induction to show that for $r \leq l \leq k$,

$$\mathbb{E}_l W_l \dots W_k U \leq \exp m_l(f_l - F_l). \quad (3.19)$$

As in (3.7), $f \leq F$ implies that $f_l \leq F_l$ and since $m_{l-1} \leq m_l$, (3.19) implies that

$$\mathbb{E}_l W_l \dots W_k U \leq \exp m_{l-1}(f_l - F_l).$$

Multiplying both sides by W_{l-1} gives

$$\mathbb{E}_l W_{l-1} W_l \dots W_k U \leq \exp m_{l-1}(f_l - F_{l-1})$$

since W_l does not depend on z_i for $i \geq l$. Taking the expectation \mathbb{E}_{l-1} and using that F_{l-1} does not depend on z_i for $i \geq l-1$ we can write

$$\mathbb{E}_{l-1} W_{l-1} W_l \dots W_k U \leq \exp(-m_{l-1} F_{l-1}) \mathbb{E}_{l-1} \exp m_{l-1} f_l = \exp m_{l-1}(f_{l-1} - F_{l-1}).$$

This finishes the proof of the induction step. For $l = r$, (3.19) implies that

$$\log \mathbb{E}_r W_r \dots W_k U \leq m(f_r - F_r). \quad (3.20)$$

Since for $l < r$, $m_l = 0$, (3.2) implies that

$$\mathcal{P}(\mathbf{m})f = f_0 = \mathbb{E}_0 f_r \text{ and } \mathcal{P}(\mathbf{m})F = F_0 = \mathbb{E}_0 F_r$$

and, therefore, taking the expectation of both sides of (3.20) proves (3.18). □

4 Guerra's interpolation.

The first step of the proof of Theorem 1 is the analogue of Guerra's interpolation method in [2]. For $1 \leq i \leq N$, we consider independent copies $(z_{i,p})_{0 \leq p \leq k}$ of the sequence $(z_p)_{0 \leq p \leq k}$ defined in (1.12) that are also independent of the randomness of the Hamiltonian $H_N(\boldsymbol{\sigma})$. Consider $\mathbf{m} = (m_p)_{0 \leq p \leq k}$ as in (1.10). We denote by \mathbb{E}_l the expectation in the r.v. $(z_{i,p})_{i \leq N, p \geq l}$. Consider the Hamiltonian

$$H_t(\boldsymbol{\sigma}) = \sqrt{t}H_N(\boldsymbol{\sigma}) + \sqrt{1-t} \sum_{i \leq N} \sigma_i \left(\sum_{0 \leq p \leq k} z_{i,p} \right). \quad (4.1)$$

Given U_N defined in (1.8), let

$$F = \log \int_{U_N} \exp H_t(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}) \quad (4.2)$$

and let

$$\varphi_N(t) = \frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{m}) F, \quad (4.3)$$

where \mathbb{E} denotes the expectation in all random variables including the randomness of the Hamiltonian $H_N(\boldsymbol{\sigma})$.

For a function $h : \Sigma^N \rightarrow \mathbb{R}$, let $\langle h \rangle_t$ denote its average with respect to the Gibbs measure with Hamiltonian $H_t(\boldsymbol{\sigma})$ in (4.1) on the set U_N , i.e.

$$\langle h \rangle_t \exp F = \int_{U_N} h(\boldsymbol{\sigma}) \exp H_t(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}). \quad (4.4)$$

(3.14) implies that the functional

$$h \rightarrow \mathbb{E}_l(W_l \dots W_k \langle h \rangle_t) \quad (4.5)$$

is a probability γ_l on U_N . We denote by $\gamma_l^{\otimes 2}$ its product on $U_N \times U_N$, and for a function $h : U_N \times U_N \rightarrow \mathbb{R}$ we set

$$\mu_l(h) = \mathbb{E}(W_1 \dots W_{l-1} \gamma_l^{\otimes 2}(h)). \quad (4.6)$$

The following Gaussian integration by parts will be commonly used below. If g is a Gaussian random variable then for a function $F : \mathbb{R} \rightarrow \mathbb{R}$ of moderate growth we have (A.40 in [8]),

$$\mathbb{E} g F(g) = \mathbb{E} g^2 \mathbb{E} F'(g). \quad (4.7)$$

This can be generalized as follows. If $\mathbf{g} = (g_1, \dots, g_n)$ is a jointly Gaussian family of random variables then for a function $F : \mathbb{R}^n \rightarrow \mathbb{R}$ of moderate growth we have (see for example, A.41 in [8]),

$$\mathbb{E} g_i F(\mathbf{g}) = \sum_{j \leq n} \mathbb{E}(g_i g_j) \mathbb{E} \frac{\partial F}{\partial g_j}(\mathbf{g}). \quad (4.8)$$

We will need a similar statement for functionals of not necessarily finite Gaussian families, for example, for a random process $H_N(\boldsymbol{\sigma})$ indexed by $\boldsymbol{\sigma}$ in a possibly infinite set Σ^N . The following is a simple consequence of (4.7).

Lemma 4 *Let $\mathbf{g} = (g(\boldsymbol{\rho}))_{\boldsymbol{\rho} \in U}$ be a Gaussian process indexed by $U \subseteq \mathbb{R}^n$ and let $F(\mathbf{g})$ be a differentiable functional on \mathbb{R}^U . Given $\boldsymbol{\sigma} \in U$, we have*

$$\mathbb{E}g(\boldsymbol{\sigma})F(\mathbf{g}) = \mathbb{E}\frac{\delta F}{\delta \mathbf{g}}[\mathbb{E}g(\boldsymbol{\sigma})g(\boldsymbol{\rho})] \quad (4.9)$$

- the expectation of the variational derivative of F in the direction $h(\boldsymbol{\rho}) = \mathbb{E}g(\boldsymbol{\sigma})g(\boldsymbol{\rho})$.

Proof. Consider a process $\mathbf{g}' = (g'(\boldsymbol{\rho}))_{\boldsymbol{\rho} \in U}$ defined by

$$g'(\boldsymbol{\rho}) = g(\boldsymbol{\rho}) - g(\boldsymbol{\sigma})\frac{\mathbb{E}g(\boldsymbol{\rho})g(\boldsymbol{\sigma})}{\mathbb{E}g(\boldsymbol{\sigma})^2},$$

which is, obviously, independent of the r.v. $g(\boldsymbol{\sigma})$. If we fix \mathbf{g}' and denote by \mathbb{E}' the expectation with respect to $g(\boldsymbol{\sigma})$ then, using (4.7) with $g = g(\boldsymbol{\sigma})$ gives

$$\mathbb{E}'g(\boldsymbol{\sigma})F(\mathbf{g}) = \mathbb{E}'g(\boldsymbol{\sigma})F\left(g'(\boldsymbol{\rho}) + g(\boldsymbol{\sigma})\frac{\mathbb{E}g(\boldsymbol{\rho})g(\boldsymbol{\sigma})}{\mathbb{E}g(\boldsymbol{\sigma})^2}\right) = \mathbb{E}'\frac{\delta F}{\delta \mathbf{g}}[\mathbb{E}g(\boldsymbol{\sigma})g(\boldsymbol{\rho})].$$

Taking the expectation in \mathbf{g}' proves (4.9). □

We are ready to prove the main result of this section. The proof will clarify why we first compute the free energy of the set of configurations with constrained self-overlap $R_{1,1} \in [u - \varepsilon_N, u + \varepsilon_N]$.

Theorem 4 (*Guerra's interpolation*). *For $t \in [0, 1]$ we have*

$$\begin{aligned} \varphi'_N(t) &= -\frac{1}{2} \sum_{1 \leq l \leq k} m_l(\theta(q_{l+1}) - \theta(q_l)) \\ &\quad - \frac{1}{2} \sum_{1 \leq l \leq k} (m_l - m_{l-1})\mu_l(\xi(R_{1,2}) - R_{1,2}\xi'(q_l) + \theta(q_l)) + \mathcal{R}, \end{aligned} \quad (4.10)$$

where $|\mathcal{R}| \leq c(N) + L\varepsilon_N$.

Proof. The proof of this theorem repeats the proof of the main result in [2] (see also Theorem 2.1 in [11]) with some necessary modifications. We will give the detailed proof in order to demonstrate how Lemma 4 replaces (4.8) and to show that the constraint $q_{k+1} = u$ in (1.11), in some sense, matches the constraint on the self overlap $R_{1,1} \in [u - \varepsilon_N, u + \varepsilon_N]$ in the definition of local free energy (1.8), (1.9). In particular, we will see in (4.21) below that such choice of parameters allows us to get rid of a certain term in the derivative $\varphi'_N(t)$ that, otherwise, would be problematic to control. First of all, (3.15) implies that

$$\varphi'_N(t) = \frac{1}{N} \mathbb{E} W_1 \dots W_k \frac{\partial F}{\partial t}.$$

Using the fact that $m_k = 1$, one can write

$$\begin{aligned} W_1 \dots W_k &= \exp \sum_{1 \leq l \leq k} m_l (F_{l+1} - F_l) \\ &= \exp \left(F + \sum_{1 \leq l \leq k} (m_{l-1} - m_l) F_l \right) = T \exp F, \end{aligned} \quad (4.11)$$

where $T = T_1 \dots T_k$ and $T_l = \exp(m_{l-1} - m_l) F_l$. Using (4.2) we can write

$$\varphi'_N(t) = \frac{1}{N} \mathbb{E} T \exp F \frac{\partial F}{\partial t} = \text{I} - \text{II},$$

where

$$\text{I} = \frac{1}{2N\sqrt{t}} \int_{U_N} \mathbb{E} T H_N(\boldsymbol{\sigma}) \exp H_t(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}). \quad (4.12)$$

and

$$\text{II} = \frac{1}{2N\sqrt{1-t}} \int_{U_N} \mathbb{E} T \sum_{i \leq N} \sigma_i \sum_{p \leq k} z_{i,p} \exp H_t(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}). \quad (4.13)$$

To compute I, we will use (4.9) for the family $\mathbf{g} = (H_N(\boldsymbol{\rho}))_{\boldsymbol{\rho} \in U_N}$ and we will think of each factor T_l in T as the functional of \mathbf{g} . Let us denote

$$\zeta(\boldsymbol{\sigma}, \boldsymbol{\rho}) = \frac{1}{N} \mathbb{E} H_N(\boldsymbol{\sigma}) H_N(\boldsymbol{\rho}). \quad (4.14)$$

Then (4.9) and (4.14) imply

$$\begin{aligned} \text{I} &= \frac{1}{2\sqrt{t}} \int_{U_N} \mathbb{E} T \frac{\partial \exp H_t(\boldsymbol{\sigma})}{\partial H_N(\boldsymbol{\sigma})} \zeta(\boldsymbol{\sigma}, \boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}) \\ &+ \frac{1}{2\sqrt{t}} \sum_{l \leq k} \int_{U_N} \mathbb{E} T \exp H_t(\boldsymbol{\sigma}) \frac{1}{T_l} \frac{\delta T_l}{\delta \mathbf{g}} [\zeta(\boldsymbol{\sigma}, \boldsymbol{\rho})] d\nu(\boldsymbol{\sigma}). \end{aligned} \quad (4.15)$$

First of all,

$$\frac{\partial \exp H_t(\boldsymbol{\sigma})}{\partial H_N(\boldsymbol{\sigma})} = \sqrt{t} \exp H_t(\boldsymbol{\sigma}),$$

and, therefore, the first line in (4.15) can be written as

$$\frac{1}{2} \int_{U_N} \mathbb{E} T \exp H_t(\boldsymbol{\sigma}) \zeta(\boldsymbol{\sigma}, \boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}) = \frac{1}{2} \mathbb{E} W_1 \dots W_k \langle \zeta(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \rangle_t. \quad (4.16)$$

Using the definition of T_l we can write

$$\frac{1}{T_l} \frac{\delta T_l}{\delta \mathbf{g}} [\zeta(\boldsymbol{\sigma}, \boldsymbol{\rho})] = (m_{l-1} - m_l) \frac{\delta F_l}{\delta \mathbf{g}} [\zeta(\boldsymbol{\sigma}, \boldsymbol{\rho})] = (m_{l-1} - m_l) \mathbb{E}_l W_l \dots W_k \frac{\delta F}{\delta \mathbf{g}} [\zeta(\boldsymbol{\sigma}, \boldsymbol{\rho})],$$

where we used (3.16). We have

$$\frac{\delta F}{\delta \mathbf{g}} [\zeta(\boldsymbol{\sigma}, \boldsymbol{\rho})] = \sqrt{t} \exp(-F) \int_{U_N} \zeta(\boldsymbol{\sigma}, \boldsymbol{\rho}) \exp H_t(\boldsymbol{\rho}) d\nu(\boldsymbol{\rho}) = \sqrt{t} \langle \zeta(\boldsymbol{\sigma}, \boldsymbol{\rho}) \rangle'_t,$$

where $\langle \cdot \rangle'_t$ denotes the Gibbs average with respect to $\boldsymbol{\rho}$ for a fixed $\boldsymbol{\sigma}$. Therefore, for a fixed $\boldsymbol{\sigma}$ we get

$$\frac{1}{T_l} \frac{\delta T_l}{\delta \mathbf{g}} [\zeta(\boldsymbol{\sigma}, \boldsymbol{\rho})] = \sqrt{t}(m_{l-1} - m_l) \mathbb{E}_l W_l \dots W_k \langle \zeta(\boldsymbol{\sigma}, \boldsymbol{\rho}) \rangle'_t = \sqrt{t}(m_{l-1} - m_l) \gamma_l(\zeta(\boldsymbol{\sigma}, \boldsymbol{\rho})),$$

where γ_l was defined in (4.5). Hence, the second line in (4.15) is equal to

$$\begin{aligned} & \frac{1}{2} \sum_{l \leq k} (m_{l-1} - m_l) \mathbb{E} W_1 \dots W_k \exp(-F) \int_{U_N} \gamma_l(\zeta(\boldsymbol{\sigma}, \boldsymbol{\rho})) \exp H_t(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}) \\ &= \frac{1}{2} \sum_{l \leq k} (m_{l-1} - m_l) \mathbb{E} W_1 \dots W_{l-1} \gamma_l^{\otimes 2}(\zeta(\boldsymbol{\sigma}, \boldsymbol{\rho})) = \frac{1}{2} \sum_{l \leq k} (m_{l-1} - m_l) \mu_l(\zeta(\boldsymbol{\sigma}, \boldsymbol{\rho})), \end{aligned} \quad (4.17)$$

where μ_l was defined in (4.6). Combining (4.16) and (4.17) we get

$$\text{I} = \frac{1}{2} \mathbb{E} W_1 \dots W_k \langle \zeta(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \rangle_t + \frac{1}{2} \sum_{l \leq k} (m_{l-1} - m_l) \mu_l(\zeta(\boldsymbol{\sigma}, \boldsymbol{\rho})). \quad (4.18)$$

By (1.1), for any $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2 \in \Sigma^N$ we have

$$|\zeta(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) - \xi(R_{1,2})| \leq c(N)$$

and, therefore, (4.18) implies that

$$\text{I} = \frac{1}{2} \mathbb{E} W_1 \dots W_k \langle \xi(R_{1,1}) \rangle_t + \frac{1}{2} \sum_{l \leq k} (m_{l-1} - m_l) \mu_l(\xi(R_{1,2})) + \mathcal{R}, \quad (4.19)$$

where $|\mathcal{R}| \leq c(N)$. The computation of II is very similar, one only needs to note that F_l does not depend on $z_{i,p}$ for $l \leq p$. We have

$$\text{II} = \frac{1}{2} \xi'(q_{k+1}) \mathbb{E} W_1 \dots W_k \langle R_{1,1} \rangle_t + \frac{1}{2} \sum_{l \leq k} (m_{l-1} - m_l) \xi'(q_l) \mu_l(R_{1,2}). \quad (4.20)$$

Combining (4.19) and (4.20) and rearranging terms, it is easy to see that

$$\begin{aligned} \varphi'_N(t) &= \frac{1}{2} \mathbb{E} W_1 \dots W_k \langle \xi(R_{1,1}) - R_{1,1} \xi'(q_{k+1}) + \theta(q_{k+1}) \rangle_t - \frac{1}{2} \sum_{l \leq k} m_l (\theta(q_{l+1}) - \theta(q_l)) \\ &\quad - \frac{1}{2} \sum_{1 \leq l \leq k} (m_l - m_{l-1}) \mu_l(\xi(R_{1,2}) - R_{1,2} \xi'(q_l) + \theta(q_l)) + \mathcal{R}. \end{aligned} \quad (4.21)$$

It now suffices to notice that $\langle \cdot \rangle_t$ is restricted to the set U_N and $q_{k+1} = u$ by (1.11), which implies that

$$0 \leq \xi(R_{1,1}) - R_{1,1} \xi'(u) + \theta(u) \leq L |R_{1,1} - u| \leq L \varepsilon_N.$$

Since $\mathbb{E} W_1 \dots W_k = 1$, this finishes the proof of Theorem 4. It is to control the first term on the right hand side of (4.21) that we impose the constraint on self-overlap. In the classical Sherrington-Kirkpatrick model this problem did not occur because $R_{1,1}$ was always 1.

□

5 Removing the constraint on the self-overlap.

The main goal of this section is to compute $\lim_{N \rightarrow \infty} \varphi_N(0)$. The convexity of ξ implies that $\xi(a) - a\xi'(b) + \theta(b) \geq 0$ for any $a, b \in \mathbb{R}$ and, therefore, (4.10) implies

$$\varphi'_N(t) \leq -\frac{1}{2} \sum_{1 \leq l \leq k} m_l(\theta(q_{l+1}) - \theta(q_l)) + \mathcal{R}$$

and, hence,

$$F_N(u, \varepsilon_N) = \varphi_N(1) \leq \varphi_N(0) - \frac{1}{2} \sum_{1 \leq l \leq k} m_l(\theta(q_{l+1}) - \theta(q_l)) + \mathcal{R}. \quad (5.1)$$

In the SK model $\varphi_N(0)$ was very easy to compute and, in fact, it was independent of N because at the end of Guerra's interpolation the spins became decoupled. The situation is different here because of the constraint (1.8). Let us recall that for $t = 0$,

$$\varphi_N(0) = \frac{1}{N} \mathcal{P}(\mathbf{m}) F \quad \text{where} \quad F = \log \int_{U_N} \exp \sum_{i \leq N} \sigma_i \left(\sum_{0 \leq p \leq k} z_{i,p} \right) d\nu(\boldsymbol{\sigma}).$$

Removing the constraint U_N constitutes a large deviation problem that will be addressed in Theorem 5 below. First of all, let us give an easy upper bound on $\varphi_N(0)$. Since by (1.7) the self-overlap $R_{1,1} \in [d, D]$, for $\boldsymbol{\sigma} \in U_N$ we have $R_{1,1} \in [d, D] \cap [u - \varepsilon_N, u + \varepsilon_N]$ and, therefore,

$$-\lambda R_{1,1} \leq -\lambda u_N(\lambda), \quad (5.2)$$

where

$$u_N(\lambda) = \max(d, u - \varepsilon_N) \text{ for } \lambda \geq 0 \text{ and } u_N(\lambda) = \min(D, u + \varepsilon_N) \text{ for } \lambda < 0. \quad (5.3)$$

Using this, we can bound F as follows,

$$\begin{aligned} F &= \log \int_{U_N} \exp \left(\sum_{i \leq N} \sigma_i \sum_{0 \leq p \leq k} z_{p,i} \right) d\nu(\boldsymbol{\sigma}) \\ &\leq -N\lambda u_N(\lambda) + \log \int_{U_N} \exp \left(\sum_{i \leq N} \sigma_i \sum_{0 \leq p \leq k} z_{p,i} + \lambda \sum_{i \leq N} \sigma_i^2 \right) d\nu(\boldsymbol{\sigma}) \\ &\leq -N\lambda u_N(\lambda) + \log \int_{\Sigma^N} \exp \left(\sum_{i \leq N} \sigma_i \sum_{0 \leq p \leq k} z_{p,i} + \lambda \sum_{i \leq N} \sigma_i^2 \right) d\nu(\boldsymbol{\sigma}) \\ &= -N\lambda u_N(\lambda) + \sum_{i \leq N} X_{k+1,i} \end{aligned}$$

where

$$X_{k+1,i} = \log \int_{\Sigma} \left(\sigma \sum_{0 \leq p \leq k} z_{p,i} + \lambda \sigma^2 \right) d\nu(\sigma) \quad (5.4)$$

are independent copies of X_{k+1} defined in (1.13). Using (3.5) and (3.7),

$$\varphi_N(0) = \frac{1}{N} \mathcal{P}(\mathbf{m}) F \leq -\lambda u_N(\lambda) + \mathcal{P}(\mathbf{m}) X_{k+1} = -\lambda u_N(\lambda) + X_0.$$

The arbitrary choice of λ here implies that

$$\varphi_N(0) \leq \inf_{\lambda} (-\lambda u_N(\lambda) + X_0). \quad (5.5)$$

Combining (5.1) and (5.5) we get

$$F_N(u, \varepsilon_N) \leq \inf \left(-\lambda u_N(\lambda) + X_0(\mathbf{m}, \mathbf{q}, \lambda) - \frac{1}{2} \sum_{1 \leq l \leq k} m_l (\theta(q_{l+1}) - \theta(q_l)) \right) + \mathcal{R}, \quad (5.6)$$

where the infimum is over all choices of parameters $k, \mathbf{m}, \mathbf{q}$ and λ . The bound (5.6) is the analogue of Guerra's replica symmetry breaking bound in [2]. If instead of $u_N(\lambda)$ we had u in (5.6) then the infimum would be equal to $\mathcal{P}(\xi, u)$ which would prove the upper bound in Theorem 1. We will now show that for

$$d < u < D \quad (5.7)$$

this infimum is not changed much by replacing $u_N(\lambda)$ with u . We will need the following.

Lemma 5 *There exists a function $a(\lambda)$ such that for any $k, \mathbf{m}, \mathbf{q}$,*

$$a(\lambda) \lambda \leq X_0(\mathbf{m}, \mathbf{q}, \lambda) \quad (5.8)$$

and such that

$$\lim_{\lambda \rightarrow +\infty} a(\lambda) = D \text{ and } \lim_{\lambda \rightarrow -\infty} a(\lambda) = d. \quad (5.9)$$

Proof. Indeed, if in the recursive construction (1.14) one takes all $m_l = 0$ then Hölder's inequality yields that for any sequence \mathbf{m}

$$\mathbb{E} \log \int_{\Sigma} \exp \left(\sigma \sum_{0 \leq p \leq k} z_p + \lambda \sigma^2 \right) d\nu(\sigma) \leq X_0(\mathbf{m}, \mathbf{q}, \lambda) \quad (5.10)$$

Since the function

$$x \rightarrow \log \int_{\Sigma} \exp \left(\sigma x + \lambda \sigma^2 \right) d\nu(\sigma)$$

is convex by Hölder's inequality, (5.10) and Jensen's inequality imply that

$$\log \int_{\Sigma} \exp \lambda \sigma^2 d\nu(\sigma) \leq X_0(\mathbf{m}, \mathbf{q}, \lambda). \quad (5.11)$$

It is clear that (1.7) implies that the left hand side of (5.11) is asymptotically equivalent to $D\lambda$ for $\lambda \rightarrow +\infty$ and to $d\lambda$ for $\lambda \rightarrow -\infty$ and this proves Lemma 5. □

Next we will show that the estimate (5.8) and equation (5.6) imply the upper bound in Theorem 1.

Lemma 6 For $d < u < D$ we have

$$\limsup_{N \rightarrow \infty} F_N(u, \varepsilon_N) \leq \mathcal{P}(\xi, u). \quad (5.12)$$

Proof. Lemma 5 implies that

$$-\lambda u_N(\lambda) + X_0(\mathbf{m}, \mathbf{q}, \lambda) \geq \lambda(a(\lambda) - u_N(\lambda)). \quad (5.13)$$

For $d < u < D$, the definition (5.3) implies that for any λ , $\lim_{N \rightarrow \infty} u_N(\lambda) = u$. Combining this with (5.9) yields that for N large enough the right hand side of (5.13) goes to infinity as $\lambda \rightarrow \pm\infty$ and, thus, the infimum in (5.6) is achieved for $|\lambda| \leq \Lambda$ where Λ is a large enough constant independent of $k, \mathbf{m}, \mathbf{q}$. This means that in (5.6) restricting minimization over λ to the set $\{|\lambda| \leq \Lambda\}$ does not change the infimum and, therefore,

$$\begin{aligned} F_N(u, \varepsilon_N) &\leq \inf \left(-\lambda u + X_0(\mathbf{m}, \mathbf{q}, \lambda) - \frac{1}{2} \sum_{1 \leq l \leq k} m_l (\theta(q_{l+1}) - \theta(q_l)) \right) + \Lambda |u_N(\lambda) - u| + \mathcal{R}, \\ &= \mathcal{P}(\xi, u) + \Lambda |u_N(\lambda) - u| + \mathcal{R}. \end{aligned}$$

and this finishes the proof. \square

In the rest of the section we will show that the bound in (5.5) is exact in the limit.

Theorem 5 For any $d < u < D$, if the sequence ε_N goes to zero slowly enough then

$$\lim_{N \rightarrow \infty} \varphi_N(0) = \varphi_0(\mathbf{m}, \mathbf{q}) := \inf_{\lambda} (-\lambda u + X_0(\mathbf{m}, \mathbf{q}, \lambda)). \quad (5.14)$$

Proof. Given a measurable set $A \subseteq [d, D]$ and $\lambda \in \mathbb{R}$, we define

$$F(A, \lambda) = \log \int_{\{R_{1,1} \in A\}} \exp \left(\sum_{i=1}^N \sigma_i \sum_{0 \leq p \leq k} z_{i,p} + \lambda \sum_{i=1}^N \sigma_i^2 \right) d\nu(\boldsymbol{\sigma}) \quad (5.15)$$

and

$$\Phi(A, \lambda) = \frac{1}{N} \mathcal{P}(\mathbf{m}) F(A, \lambda). \quad (5.16)$$

When $A = [d, D]$, the set $\{R_{1,1} \in A\} = \Sigma^N$ and, therefore,

$$F([d, D], \lambda) = \sum_{i \leq N} X_{k+1,i}$$

where $X_{k+1,i}$ were defined in (5.4). Using (3.6),

$$\Phi([d, D], \lambda) = \mathcal{P}(\mathbf{m}) X_{k+1} = X_0(\lambda). \quad (5.17)$$

Here we made the dependence of X_0 on λ explicit while keeping the dependence on other parameters implicit. For simplicity of notations we will write

$$F(\lambda) := F([d, D], \lambda) \text{ and } \Phi(\lambda) := \Phi([d, D], \lambda) = X_0(\lambda).$$

In these notations Theorem 5 states that there exists a sequence $\varepsilon_N \rightarrow 0$ such that

$$\lim_{N \rightarrow \infty} \Phi(U_N, 0) = \inf_{\lambda} (-\lambda u + \Phi(\lambda)) = \inf_{\lambda} (-\lambda u + X_0(\lambda)). \quad (5.18)$$

Let us define $\lambda(u)$ to be the point where the infimum in (5.18) is achieved, i.e.

$$X_0(\lambda(u)) - \lambda(u)u = \inf_{\lambda} (X_0(\lambda) - \lambda u). \quad (5.19)$$

The infimum is, indeed, achieved because of the following argument. A function $-\lambda u + X_0(\lambda)$ is convex in λ by Hölder's inequality. Lemma 5 implies that $X_0(\lambda) \geq a(\lambda)\lambda$ for some function $a(\lambda)$ such that $\lim_{\lambda \rightarrow +\infty} a(\lambda) = D$ and $\lim_{\lambda \rightarrow -\infty} a(\lambda) = d$. Hence, for $d < u < D$, the convex function $-\lambda u + X_0(\lambda) \rightarrow +\infty$ as $\lambda \rightarrow \pm\infty$ and, therefore, it has a unique minimum. The critical point condition for $\lambda(u)$ is

$$\frac{\partial X_0}{\partial \lambda}(\lambda(u)) = u. \quad (5.20)$$

Consider a fixed small enough $\varepsilon > 0$ such that $d < u - \varepsilon$ and $u + \varepsilon < D$ and let $U_\varepsilon = [u - \varepsilon, u + \varepsilon]$. Let us analyze $\Phi(U_\varepsilon, \lambda(u))$. Consider a set V equal to either $[d, u - \varepsilon]$ or $[u + \varepsilon, D]$ and note that

$$[d, D] = U_\varepsilon \cup [d, u - \varepsilon] \cup [u + \varepsilon, D].$$

We will start by proving an upper bound on $\Phi(V, \lambda(u))$. We will only consider the case of $V = [u + \varepsilon, D]$ since the case $V = [d, u - \varepsilon]$ can be treated similarly. Since

$$-\gamma R_{1,1} \leq -\gamma(u + \varepsilon) \quad \text{for } R_{1,1} \in V \text{ and } \gamma \geq 0,$$

we get that for $\gamma \geq 0$,

$$F(V, \lambda(u)) \leq -N\gamma(u + \varepsilon) + F(\lambda(u) + \gamma).$$

Using (3.5) and (3.7) we get

$$\Phi(V, \lambda(u)) \leq U(\gamma) := -\gamma(u + \varepsilon) + \Phi(\lambda(u) + \gamma) = -\gamma(u + \varepsilon) + X_0(\lambda(u) + \gamma). \quad (5.21)$$

Setting $\gamma = 0$ gives

$$U(0) = \Phi(\lambda(u)) = X_0(\lambda(u)).$$

Next, the right derivative of $U(\gamma)$ at zero is

$$\left. \frac{\partial U}{\partial \gamma} \right|_{\gamma=0^+} = -(u + \varepsilon) + \frac{\partial X_0}{\partial \lambda}(\lambda(u)) = -(u + \varepsilon) + u = -\varepsilon$$

using (5.20). Finally, it follows from a tedious but straightforward computation which we will omit here that

$$\left| \frac{\partial^2 U}{\partial \gamma^2} \right| \leq L$$

for some constant L that depends only on the parameters of the model ξ and ν . Therefore, minimizing over $\gamma \geq 0$ in the right hand side of (5.21) gives

$$\Phi(V, \lambda(u)) \leq \Phi(\lambda(u)) - \frac{\varepsilon^2}{L}. \quad (5.22)$$

The same bound holds for $V = [d, u - \varepsilon]$. For $l \geq 1$, let

$$W_l = \exp m_l(F_{l+1}(\lambda(u)) - F_l(\lambda(u)))$$

be defined by (3.13) with $F = F(\lambda(u))$. Given a set $A \subseteq [d, D]$ let $\langle I(R_{1,1} \in A) \rangle$ denote the Gibbs average defined by

$$\langle I(R_{1,1} \in A) \rangle = \exp(F(A, \lambda(u)) - F(\lambda(u))).$$

The following Proposition is the crucial step in the proof of Theorem 5. This type of computation was invented by Talagrand in [11] in order to control the remainder terms in Guerra's interpolation and we will use this argument with the same purpose later in the paper as well.

Proposition 1 *Assume that for $A \subseteq [d, D]$ and for some $\varepsilon' > 0$ we have*

$$\Phi(A, \lambda(u)) \leq \Phi(\lambda(u)) - \varepsilon'. \quad (5.23)$$

Then,

$$\mathbb{E} W_1 \dots W_k \langle I(R_{1,1} \in A) \rangle \leq L \exp\left(-\frac{N}{L}\right), \quad (5.24)$$

where L does not depend on N .

Proof. The proof is based on the property of the Parisi functional described in Lemma 3. For simplicity of notations let us assume that $m_1 > 0$. The case when several elements of the sequence \mathbf{m} are zeroes can be handled in exactly the same way. Let

$$f = F(A, \lambda(u)) \text{ and } F = F(\lambda(u))$$

so that the condition $f \leq F$ of Lemma 3 is satisfied. Lemma 3 and (5.23) imply that

$$\begin{aligned} \mathbb{E} \log \mathbb{E}_1 W_1 \dots W_k \exp(f - F) &\leq m_1(\mathcal{P}(\mathbf{m})f - \mathcal{P}(\mathbf{m})F) \\ &= Nm_1(\Phi(A, \lambda(u)) - \Phi(\lambda(u))) \leq -Nm_1\varepsilon'. \end{aligned}$$

Let us consider a function

$$\phi(Z) = \log \mathbb{E}_1 W_1 \dots W_k \exp(f - F) \text{ where } Z = (z_{i,0})_{i \leq N}. \quad (5.25)$$

We will show that $\phi(Z)$ is a Lipschitz function of Z :

$$|\phi(Z) - \phi(Z')| \leq L\sqrt{N}|Z - Z'|. \quad (5.26)$$

First of all,

$$\sum_{i \leq N} \sigma_i(z_{i,0} - z'_{i,0}) \leq |Z - Z'| \left(\sum_{i \leq N} \sigma_i^2 \right)^{1/2} \leq \sqrt{ND} |Z - Z'|,$$

using (1.7). Definition (5.15) implies that for any set A and any λ ,

$$|F(A, \lambda)(Z) - F(A, \lambda)(Z')| \leq \sqrt{ND} |Z - Z'|,$$

where we made the dependence of $F(A, \lambda)$ on Z explicit. In particular, this holds for f and F . It is also clear from the properties (3.5) and (3.7) that iteration (3.2) in the definition of the Parisi functional preserves Lipschitz condition and, therefore,

$$|F_l(Z) - F_l(Z')| \leq \sqrt{ND} |Z - Z'| \text{ for } l \leq k.$$

Using (4.11) we can rewrite $\phi(Z)$ as

$$\phi(Z) = \log \mathbb{E}_1 \exp \left(f(Z) + \sum_{1 \leq l \leq k} (m_{l-1} - m_l) F_l(Z) \right)$$

and (5.26) is now obvious. Gaussian concentration of measure (see, for example, Theorem 2.2.4 in [8]) implies that for any $t \geq 0$,

$$\mathbb{P}(|\phi(Z) - \mathbb{E}\phi| \geq Nt) \leq 2 \exp(-Nt^2/L). \quad (5.27)$$

In particular, with probability at least $1 - 2 \exp(-N/L)$,

$$\phi \leq \mathbb{E}\phi + \frac{1}{2} N m_1 \varepsilon' \leq -\frac{1}{2} N m_1 \varepsilon'$$

and, therefore, with probability at least $1 - 2 \exp(-N/L)$,

$$\mathbb{E}_1 W_1 \dots W_k \exp(f - F) \leq \exp(-N/L). \quad (5.28)$$

Since $f \leq F$,

$$\mathbb{E}_1 W_1 \dots W_k \exp(f - F) \leq \mathbb{E}_1 W_1 \dots W_k \leq 1$$

using (3.14) which together with (5.28) implies that

$$\mathbb{E} W_1 \dots W_k \exp(f - F) \leq L \exp(-N/L).$$

□

Corollary 1 *For any $\varepsilon > 0$ we have*

$$\Phi(U_\varepsilon, \lambda(u)) \geq \Phi(\lambda(u)) - \delta_N = X_0(\lambda(u)) - \delta_N, \quad (5.29)$$

where $\lim_{N \rightarrow \infty} \delta_N = 0$.

Proof. (5.22) and (5.24) imply that for V equal to either $[d, u - \varepsilon]$ or $[u + \varepsilon, D]$ we have

$$\mathbb{E}W_1 \dots W_k \langle I(R_{1,1} \in V) \rangle \leq L \exp\left(-\frac{N}{L}\right)$$

and, therefore,

$$\mathbb{E}W_1 \dots W_k \langle I(R_{1,1} \notin U_\varepsilon) \rangle \leq L \exp\left(-\frac{N}{L}\right). \quad (5.30)$$

Suppose that (5.29) is not true which means that for some positive $\varepsilon' > 0$ we have

$$\Phi(U_\varepsilon, \lambda(u)) \leq \Phi(\lambda(u)) - \varepsilon',$$

for some arbitrarily large N . Then again (5.22) and (5.24) would imply that

$$\mathbb{E}W_1 \dots W_k \langle I(R_{1,1} \in U_\varepsilon) \rangle \leq L \exp\left(-\frac{N}{L}\right).$$

Combining with (5.30) we would get

$$1 = \mathbb{E}W_1 \dots W_k \leq L \exp\left(-\frac{N}{L}\right)$$

and we arrive at contradiction. □

In order to bound $\Phi(U_\varepsilon, \lambda(u))$ in terms of $\Phi(U_\varepsilon, 0)$, we can write

$$\begin{aligned} F(U_\varepsilon, \lambda(u)) &= \log \int_{\{R_{1,1} \in U_\varepsilon\}} \exp\left(\sum_{i=1}^N \sigma_i \sum_{0 \leq p \leq k} z_{i,p} + N\lambda(u)R_{1,1}\right) d\nu(\boldsymbol{\sigma}) \\ &\leq N\lambda(u)u + N|\lambda(u)|\varepsilon + F(U_\varepsilon, 0). \end{aligned}$$

Using (3.5) and (3.7),

$$\Phi(U_\varepsilon, \lambda(u)) \leq \Phi(U_\varepsilon, 0) + \lambda(u)u + |\lambda(u)|\varepsilon$$

and, therefore,

$$\Phi(U_\varepsilon, 0) \geq -\lambda(u)u + X_0(\lambda(u)) - |\lambda(u)|\varepsilon - \delta_N. \quad (5.31)$$

This implies that for any $\varepsilon > 0$,

$$\liminf_{N \rightarrow \infty} \Phi(U_\varepsilon, 0) \geq -\lambda(u)u + X_0(\lambda(u)) - L\varepsilon.$$

Clearly, this means that one can choose a sequence $\varepsilon_N \rightarrow 0$ such that for $U_N = U_{\varepsilon_N}$,

$$\liminf_{N \rightarrow \infty} \Phi(U_N, 0) \geq -\lambda(u)u + X_0(\lambda(u)).$$

Since a similar upper bound is obvious this finishes the proof of Theorem 5. □

6 Reduction of the main results to apriori estimates.

Now that we understood what happens at the end of Guerra's interpolation we will turn to analyzing the remainder terms in the second line of (4.10) and, in particular, the functional μ_r defined in (4.6). First of all, for a function h on $U_N \times U_N$ the definition of $\mu_r(h)$ can be written equivalently as follows. Let (z_p^j) for $j = 1, 2$ be two copies of the random vector (z_p) defined in (1.12) such that

$$z_p^1 = z_p^2 \text{ for } p < r \text{ and } z_p^1, z_p^2 \text{ are independent for } p \geq r. \quad (6.1)$$

Let $\mathbf{z}^j = (z_{i,p}^j)_{i \leq N, p \leq k}$ where $(z_{i,p}^j)_{p \leq k}$ are independent copies of the vector (z_p^j) for $i \leq N$. Let F^j be defined by (4.2) in terms of \mathbf{z}^j and let W_l^j be defined by (3.13) in terms of F^j . Let us consider the Hamiltonian

$$H_t(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) = \sum_{j=1,2} \left(\sqrt{t} H_N(\boldsymbol{\sigma}^j) + \sqrt{1-t} \sum_{i \leq N} \sigma_i^j \left(\sum_{0 \leq p \leq k} z_{i,p}^j \right) \right) \quad (6.2)$$

and define the Gibbs average $\langle h \rangle$ of h by

$$\langle h \rangle \exp(F^1 + F^2) = \int_{U_N \times U_N} h(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) \exp H_t(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) d\nu(\boldsymbol{\sigma}^1) d\nu(\boldsymbol{\sigma}^2).$$

Then, the definition (4.6) is equivalent to

$$\mu_r(h) = \mathbb{E} W_1^1 \dots W_{r-1}^1 W_r^1 W_r^2 \dots W_k^1 W_k^2 \langle h \rangle. \quad (6.3)$$

We simply decoupled the measure $\gamma_r^{\otimes 2}$ by using independent copies of $z_{i,p}$ for $p \geq r$. Using Lemma 1 we can rewrite this in a more compact way. Let $F = F^1 + F^2$ and define \mathbf{n} as in (3.10), i.e.

$$n_p = \frac{m_p}{2} \text{ for } p < r \text{ and } n_p = m_p \text{ for } p \geq r. \quad (6.4)$$

Lemma 1 then implies that for $l \geq r$

$$W_l^1 W_l^2 = \exp m_l(F_{l+1}^1 - F_l^1) \exp m_l(F_{l+1}^2 - F_l^2) = \exp n_l(F_{l+1} - F_l)$$

and for $l < r$

$$W_l^1 = \exp m_l(F_{l+1}^1 - F_l^1) = \exp n_l(F_{l+1} - F_l).$$

Therefore, if we define

$$W_l = \exp n_l(F_{l+1} - F_l),$$

(6.3) becomes

$$\mu_r(h) = \mathbb{E} W_1 \dots W_k \langle h \rangle. \quad (6.5)$$

In particular, if $h = I(A)$ is an indicator of a measurable subset $A \subseteq U_N \times U_N$ and

$$F(A) = \log \int_A \exp H_t(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) d\nu(\boldsymbol{\sigma}^1) d\nu(\boldsymbol{\sigma}^2) \quad (6.6)$$

then, since

$$F = F^1 + F^2 = \log \int_{U_N \times U_N} \exp H_t(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) d\nu(\boldsymbol{\sigma}^1) d\nu(\boldsymbol{\sigma}^2), \quad (6.7)$$

we get

$$\mu_r(I(A)) = \mathbb{E} W_1 \dots W_k \exp(f - F).$$

Lemma 3 provided the methodology to control this expression.

Lemma 7 *Let $F(A)$ and F be defined by (6.6) and (6.7). If for some $\varepsilon > 0$ we have*

$$\frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F(A) \leq 2\varphi_N(t) - \varepsilon \quad (6.8)$$

then for some constant K independent of N and the set A ,

$$\mu_r(I(A)) \leq K \exp(-N/K). \quad (6.9)$$

Proof. Using (3.12) and (4.3), we can write

$$\frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F = \frac{2}{N} \mathbb{E} \mathcal{P}(\mathbf{m}) F^1 = 2\varphi_N(t),$$

so that (6.8) can be written as

$$\mathbb{E} \mathcal{P}(\mathbf{n}) F(A) - \mathbb{E} \mathcal{P}(\mathbf{n}) F \leq -N\varepsilon, \quad (6.10)$$

which is the type of condition used in Lemma 3. The main idea was already explained in detail in the proof of Proposition 3.5. However, the function ϕ that was defined in (5.25) was a function of the finite Gaussian vector Z in \mathbb{R}^N with independent coordinates which allowed us to use the classical Gaussian concentration of measure inequality in (5.27). Now, however, both $F(A)$ and F depend on the entire Gaussian process $H_N(\boldsymbol{\sigma})$ indexed by $\boldsymbol{\sigma} \in \Sigma^N$ and the only information that we specified about this process was the covariance operator in (1.1). Still, using the specific definition of $F(A)$ and F and (1.1) one can prove the same concentration inequality as (5.27) but it would require a tedious computation repeating the proof of (5.27) in [8]. We will actually carry out this computation in a relatively easier situation, below Lemma 8, so the idea will be clear and we will omit this computation here. The proof becomes more transparent when the Hamiltonian is expressed explicitly in terms of an i.i.d. Gaussian sequence. For example, one often considers a Hamiltonian of the type

$$H_N(\boldsymbol{\sigma}) = \sum_{p \geq 1} \frac{a_p}{N^{(p-1)/2}} \sum_{i_1, \dots, i_p} g_{i_1, \dots, i_p} \sigma_{i_1} \dots \sigma_{i_p}, \quad (6.11)$$

where $\mathbf{g} = (g_{i_1, \dots, i_p})$ is a sequence of standard Gaussian random variables independent for all $p \geq 1$ and all (i_1, \dots, i_p) . In this case,

$$\frac{1}{N} \mathbb{E} H_N(\boldsymbol{\sigma}^1) H_N(\boldsymbol{\sigma}^2) = \xi(R_{1,2}) \text{ where } \xi(x) = \sum_{p \geq 1} a_p^2 x^p.$$

The sequence $(a_p)_{p \geq 1}$ should be such that $\xi(R_{1,2})$ is well defined for all σ^1, σ^2 and, comparing with (1.7) this means that $\xi(D) < \infty$. It is easy to check that for two sequences \mathbf{g} and \mathbf{g}' we have

$$|H_N(\sigma)(\mathbf{g}) - H_N(\sigma)(\mathbf{g}')| \leq \sqrt{N\xi(D)}|\mathbf{g} - \mathbf{g}'|$$

and since inequality (5.27) is dimension independent, it applies to a sequence \mathbf{g} and the rest of the proof repeats the proof of Proposition 3.5. \square

In Lemma 6 we explained why $\mathcal{P}(\xi, u)$ is an upper bound on local free energy and the main reason was that the remainder terms in Guerra's interpolation were nonnegative. In order to show that this bound is exact in the limit we must show that these remainder terms are small along the interpolation for some choices of the parameters k, \mathbf{m} and \mathbf{q} and that $\varphi_N(t)$ can be approximated by

$$\psi(t) = \varphi_0(\mathbf{m}, \mathbf{q}) - \frac{t}{2} \sum_{1 \leq l \leq k} m_l(\theta(q_{l+1}) - \theta(q_l)), \quad (6.12)$$

where $\varphi_0(\mathbf{m}, \mathbf{q})$ was defined in (5.14). It is also clear that these parameters should approximate the infimum in the definition (1.16) of $\mathcal{P}(\xi, u)$.

Definition. We will call a vector $(k, \mathbf{m}, \mathbf{q}, \lambda)$ an ε -minimizer if

$$\mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u) \leq \mathcal{P}(\xi, u) + \varepsilon \quad \text{and} \quad (\mathbf{m}, \mathbf{q}, \lambda) \text{ is the minimizer of (1.15)}. \quad (6.13)$$

For any $v \in [-D, D]$, let us define a set

$$A(v) = \left\{ |R_{1,2} - v| \leq \frac{1}{N} \right\} \cap (U_N \times U_N). \quad (6.14)$$

The following apriori estimate will allow us to control the remainder terms in Guerra's interpolation.

Theorem 6 *For any $t_0 < 1$ there exists $\varepsilon > 0$ that depends on t_0, ξ, ν, u only such that if (6.13) holds then for $t \leq t_0$ and for large enough N ,*

$$\frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F(A(v)) \leq 2\psi(t) - \frac{(v - q_r)^2}{K} + \mathcal{R}, \quad (6.15)$$

where K is a constant independent of N, t and v and $|\mathcal{R}| \leq a_N$ for a sequence (a_N) independent of t and v and $\lim_{N \rightarrow \infty} a_N = 0$.

In the case of the replica symmetric region of Theorem 3 the condition on ε -minimizer is replaced by the condition of stability to replica symmetry breaking fluctuations defined in (2.10).

Theorem 7 *Suppose that all functions are defined in terms of parameters in (2.2) and that (2.5) and (2.10) hold. Then for any $t_0 < 1$ and for any $t \leq t_0$, for large enough N ,*

$$\frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F(A(v)) \leq 2\psi(t) - \frac{(v - q)^2}{K} + \mathcal{R}, \quad (6.16)$$

where K is a constant independent of N, t and v and $|\mathcal{R}| \leq a_N$ for a sequence (a_N) independent of t and v and $\lim_{N \rightarrow \infty} a_N = 0$.

The proof of these apriori estimates will be postponed until Appendix A. First, let us show how they imply Theorems 1 and 3.

Proof of Theorem 1. Given $t_0 < 1$ let us take $\varepsilon > 0$ as in Theorem 6 and let $(k, \mathbf{m}, \mathbf{q}, \lambda)$ be an ε -minimizer defined by (6.13). Clearly, in this case,

$$\psi(1) = \mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u) \quad \text{and} \quad |\psi(1) - \mathcal{P}(\xi, u)| \leq \varepsilon. \quad (6.17)$$

Let us take K as in (6.15) and for $\varepsilon_1 > 0$ define a set

$$\mathcal{V} = \{v \in [-D, D] : (v - q_r)^2 \geq 2K(\psi(t) - \varphi_N(t)) + 2K\varepsilon_1\}.$$

For any $v \in \mathcal{V}$, (6.15) implies that

$$\frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F(A(v)) \leq 2\psi(t) - \frac{(v - q_r)^2}{K} + \mathcal{R} \leq 2\varphi_N(t) - \varepsilon_1$$

for large enough N . Everywhere below let L denote a constant that might depend on ε_1 and K denote a constant independent of ε_1 . Applying Lemma 7, we get

$$\mu_r(I(A(v))) \leq L \exp(-N/L) \quad (6.18)$$

and the constant L here does not depend on v . Let us consider a set

$$A = \left\{ (R_{1,2} - q_r)^2 \geq 2K(\psi(t) - \varphi_N(t)) + 2K\varepsilon_1 \right\} \cap (U_N \times U_N).$$

We can choose the points $v_1, \dots, v_M \in \mathcal{V}$ with $M \leq KN$ such that $A \subseteq \bigcup_{i \leq M} A(v_i)$ and (6.18) implies that

$$\mu_r(I(A)) \leq LN \exp(-N/L) \leq L \exp(-N/L). \quad (6.19)$$

Using the definition of ψ in (6.12) and (4.10),

$$(\psi(t) - \varphi_N(t))' = \frac{1}{2} \sum_{1 \leq l \leq k} (m_l - m_{l-1}) \mu_l(\xi(R_{1,2}) - R_{1,2} \xi'(q_l) + \theta(q_l)) + \mathcal{R}. \quad (6.20)$$

Since the second derivative of ξ is bounded on $[-D, D]$,

$$\xi(R_{1,2}) - R_{1,2} \xi'(q_l) + \theta(q_l) \leq K(R_{1,2} - q_l)^2$$

and because on the complement A^c of A we have

$$(R_{1,2} - q_r)^2 \leq 2K(\psi(t) - \varphi_N(t)) + 2K\varepsilon_1,$$

(6.19) for $r = l$ implies that

$$\begin{aligned} \mu_l(\xi(R_{1,2}) - R_{1,2} \xi'(q_l) + \theta(q_l)) &\leq K \mu_l((R_{1,2} - q_l)^2) \\ &\leq K \left((\psi(t) - \varphi_N(t)) + \varepsilon_1 + \mu_l(I(A)) \right) \leq K(\psi(t) - \varphi_N(t)) + K\varepsilon_1 + L \exp(-N/L). \end{aligned}$$

(6.20) now implies that

$$(\psi(t) - \varphi_N(t))' \leq K(\psi(t) - \varphi_N(t)) + K\varepsilon_1 + L \exp(-N/L).$$

Since by Theorem 5, $\lim_{N \rightarrow \infty} \varphi_N(0) = \psi(0)$, solving this differential inequality and then letting $N \rightarrow \infty$ and $\varepsilon_1 \rightarrow 0$ implies that

$$\lim_{N \rightarrow \infty} \varphi_N(t) = \psi(t) \text{ for } t \leq t_0. \quad (6.21)$$

The derivatives $\psi'(t)$ and $\varphi'_N(t)$ are both bounded, which is apparent from (6.12) and (4.10), and we get

$$\limsup_{N \rightarrow \infty} |\varphi_N(1) - \psi(1)| \leq K(1 - t_0).$$

Using (6.17),

$$\limsup_{N \rightarrow \infty} |\varphi_N(1) - \mathcal{P}(\xi, u)| \leq K(1 - t_0) + \varepsilon$$

and letting $\varepsilon \rightarrow 0$ and $t_0 \rightarrow 1$ finishes the proof of Theorem 1. \square

Proof of Theorem 3. The proof of the the first part follows from Theorem 7 in exactly the same way as Theorem 1 follows from Theorem 6. The uniqueness of (q, λ) follows from the following simple argument. (6.21) implies that for $t \leq t_0$

$$\lim_{N \rightarrow \infty} \mu_1((R_{1,2} - q)^2) = 0$$

and since the definition of μ_1 does not depend on λ this q must be unique. Since $\mathcal{P}_1(q, \lambda)$ is convex in λ , this implies the uniqueness of λ . \square

7 Computing global free energy.

In this section we will prove Theorem 2 which will follow from Theorem 1 and Gaussian concentration of measure. Let us start by proving the following concentration inequality.

Lemma 8 *For any measurable subset $\Omega \subseteq \Sigma^N$ let us consider a r.v.*

$$X = \log \int_{\Omega} \exp H_N(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}). \quad (7.1)$$

Then, for any $t \geq 0$,

$$\mathbb{P}\left(|X - \mathbb{E}X| \geq 2\sqrt{LNt}\right) \leq 2 \exp(-t), \quad (7.2)$$

where $L = \max\{\xi(x) : x \in [d, D]\} + c(N)$.

Proof. The proof is a simple modification of Theorem 2.2.4 in [8]. Unfortunately, Lemma 8 does not fall into the framework of Theorem 2.2.4 in [8] directly, but the same argument still works if we utilize the particular definition of X and the covariance structure (1.1) of the Hamiltonian $H_N(\boldsymbol{\sigma})$.

Let H_N^1 and H_N^2 be two independent copies of the Hamiltonian H_N . For $t \in [0, 1]$, we define,

$$F_j = \log \int_{\Omega} \exp(\sqrt{t}H_N^j(\boldsymbol{\sigma}) + \sqrt{1-t}H_N(\boldsymbol{\sigma}))d\nu(\boldsymbol{\sigma}) \quad (7.3)$$

for $j = 1, 2$ and let $F = F_1 + F_2$. For $s \geq 0$, let

$$\varphi(t) = \mathbb{E} \exp s(F_2 - F_1).$$

If we define

$$H_t(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) = \sqrt{t}(H_N^1(\boldsymbol{\sigma}^1) + H_N^2(\boldsymbol{\sigma}^1)) + \sqrt{1-t}(H_N(\boldsymbol{\sigma}^1) + H_N(\boldsymbol{\sigma}^2))$$

then straightforward computation as in Theorem 4 gives

$$\varphi'(t) = -Ns^2 \mathbb{E} \exp s(F_1 - F_2) \exp(-F) \int_{\Omega^2} \zeta(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) \exp H_t(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) d\nu(\boldsymbol{\sigma}^1) d\nu(\boldsymbol{\sigma}^2), \quad (7.4)$$

where ζ was defined in (4.14). Since $|\zeta(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2)| \leq L$, we get

$$\varphi'(t) \leq LNs^2 \mathbb{E} \exp s(F_1 - F_2) \exp(-F) \int_{\Omega^2} \exp H_t d\nu d\nu = LNs^2 \varphi(t). \quad (7.5)$$

By construction $\varphi(0) = 1$, so that (7.5) implies that $\varphi(1) \leq \exp NLS^2$. On the other hand, by construction, $\varphi(1) = \mathbb{E} \exp s(X - X')$, where X is defined in (7.1) and X' is an independent copy of X and, thus,

$$\mathbb{E} \exp s(X - X') \leq \exp NLS^2.$$

By Jensen's inequality, this implies that, $\mathbb{E} \exp s(X - \mathbb{E}X) \leq \exp NLS^2$ and using Markov's inequality we get that for $t > 0$,

$$\mathbb{P}\left(X - \mathbb{E}X \geq t\right) \leq \inf_{s \geq 0} \exp(NLS^2 - st) = \exp\left(-\frac{t^2}{4NL}\right).$$

Obviously, a similar inequality can be written for $\mathbb{E}X - X$ and, therefore,

$$\mathbb{P}\left(|X - \mathbb{E}X| \geq t\right) \leq 2 \exp\left(-\frac{t^2}{4NL}\right).$$

This is equivalent to (7.2). □

The proof of Theorem 2 will be based on the concentration inequality of Lemma 8 and the following result.

Theorem 8 For $\varepsilon_N = N^{-1}$ there exists a set $\{u_1, \dots, u_M\} \subseteq [d, D]$ of cardinality $M \leq LN$ such that

$$[d, D] \subseteq \bigcup_{i \leq M} [u_i - \varepsilon_N, u_i + \varepsilon_N] \quad (7.6)$$

and such that for all $i \leq M$,

$$F_N(u_i, \varepsilon_N) \leq \sup_{d \leq u \leq D} \mathcal{P}(\xi, u) + \mathcal{R}, \quad (7.7)$$

where $|\mathcal{R}| \leq c_N + L\varepsilon_N$.

We will first show that Theorem 2 follows from Theorem 1 and Theorem 8.

Proof of Theorem 2. It follows from the definitions that for any $d \leq u \leq D$ and any sequence (ε_N) , $F_N(u, \varepsilon_N) \leq F_N$ and, therefore, Theorem 1 implies

$$\liminf_{N \rightarrow \infty} F_N \geq \sup_{d \leq u \leq D} \mathcal{P}(\xi, u).$$

To prove Theorem 2 we need to show that

$$\limsup_{N \rightarrow \infty} F_N \leq \sup_{d \leq u \leq D} \mathcal{P}(\xi, u).$$

Suppose not. Then for some $\varepsilon > 0$,

$$\limsup_{N \rightarrow \infty} F_N > \sup_{d \leq u \leq D} \mathcal{P}(\xi, u) + \varepsilon.$$

To simplify the notations, instead of considering a subsequence of N we simply assume that for N large enough we have

$$F_N \geq \sup_{d \leq u \leq D} \mathcal{P}(\xi, u) + \varepsilon. \quad (7.8)$$

On the other hand, let $\varepsilon_N = N^{-1}$ and consider a set of points $\{u_1, \dots, u_M\}$ as in Theorem 8, so that for N large enough for all $i \leq M$,

$$F_N(u_i, \varepsilon_N) \leq \sup_{d \leq u \leq D} \mathcal{P}(\xi, u) + \frac{\varepsilon}{2}. \quad (7.9)$$

Lemma 8 implies that for any $t > 0$ with probability at least $1 - LN \exp(-Nt^2/4L)$ we have

$$F_N \leq \frac{1}{N} \log Z_N + t \quad \text{and} \quad \frac{1}{N} \log Z_N(u_i, \varepsilon_N) \leq F_N(u_i, \varepsilon_N) + t \quad (7.10)$$

for all $i \leq M$. Therefore, (7.8), (7.9) and (7.10) imply that for N large enough with probability at least $1 - LN \exp(-Nt^2/4L)$ we have

$$\frac{1}{N} \log Z_N(u_i, \varepsilon_N) \leq F_N(u, \varepsilon_N) + t \leq F_N + t - \frac{\varepsilon}{2} \leq \frac{1}{N} \log Z_N + 2t - \frac{\varepsilon}{2}.$$

Taking $t = \varepsilon/K$ we get that for N large enough with probability at least $1 - LN \exp(-N\varepsilon^2/K)$ we have

$$\forall i \leq M, \quad \log Z_N(u_i, \varepsilon_N) \leq \log Z_N - \frac{N\varepsilon}{K}.$$

This yields that for $i \leq M$ the Gibbs measure

$$G_N(\{R_{1,1} \in [u_i - \varepsilon_N, u_i + \varepsilon_N]\}) \leq \exp\left(-\frac{N\varepsilon}{K}\right)$$

and, therefore, by (7.6),

$$1 = G_N(\Sigma^N) = G_N(\{R_{1,1} \in [d, D]\}) \leq LN \exp\left(-\frac{N\varepsilon}{K}\right),$$

which is impossible for large N . □

The statement of Theorem 8 is intuitively obvious considering (5.6). We, basically, need to show that the term $u_N(\lambda)$ in (5.6) can be substituted by u in a controlled manner. This will be based on three technical lemmas. To formulate the first lemma, it will be convenient to think of the pair (\mathbf{m}, \mathbf{q}) in terms of the function $m = m(q)$ defined in (A.32).

The following Lemma is the analogue of a well known continuity property of $X_0(\mathbf{m}, \mathbf{q}, \lambda)$ with respect to the functional order parameter $m(q)$ in the SK model (the statement can be found in [2] and the proof is given in [14]) and since its proof is exactly the same we will not reproduce it here.

Lemma 9 *For any λ and for any functions $m(q)$ and $m'(q)$ defined by (A.32) and corresponding to pairs (\mathbf{m}, \mathbf{q}) and $(\mathbf{m}', \mathbf{q}')$ we have*

$$|X_0(\mathbf{m}, \mathbf{q}, \lambda) - X_0(\mathbf{m}', \mathbf{q}', \lambda)| \leq L \int_0^D |m(q) - m'(q)| dq, \quad (7.11)$$

where the constant L depends on ξ and D only.

Next, we will describe several properties of the function $\mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u)$ defined in (1.15). For any k , any vectors \mathbf{m}, \mathbf{q} and any $d \leq u \leq D$ let $\lambda(u)$ be defined by (5.19).

Lemma 10 *For any $\delta > 0$ there exists a constant $\Lambda(\delta)$ such that for any vectors \mathbf{m}, \mathbf{q} and any $u \in [d + \delta, D - \delta]$ we have*

$$|\lambda(u)| \leq \Lambda(\delta). \quad (7.12)$$

Proof. For a fixed \mathbf{m} and \mathbf{q} minimizing $\mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u)$ over λ is equivalent to minimizing $-\lambda u + X_0(\mathbf{m}, \mathbf{q}, \lambda)$ over λ . By Lemma 5,

$$-\lambda u + X_0(\mathbf{m}, \mathbf{q}, \lambda) \geq \lambda(a(\lambda) - u),$$

where $a(\lambda)$ satisfies (5.9). Therefore, for any $\delta > 0$ the right hand side goes to infinity uniformly over $d + \delta \leq u \leq D - \delta$ and, thus, the left hand side goes to infinity uniformly over \mathbf{m}, \mathbf{q} and $d + \delta \leq u \leq D - \delta$. This, obviously, implies that there exists $\Lambda(\delta)$ such that (7.12) holds. □

Lemma 11 *There exists $\delta > 0$ such that for all \mathbf{m} and \mathbf{q} we have*

$$\lambda(u) < 0 \text{ for } u \in [d, d + \delta), \quad \lambda(u) > 0 \text{ for } u \in (D - \delta, D]. \quad (7.13)$$

Proof. Using (3.15) and (5.20),

$$\frac{\partial X_0}{\partial \lambda} = \mathbb{E} W_1 \dots W_k \langle \sigma^2 \rangle = u, \quad (7.14)$$

where for a function $h(\sigma)$, $\langle h \rangle$ is defined by,

$$\langle h \rangle \exp X_{k+1} = \int_{\Sigma} h(\sigma) \exp \left(\sigma \sum_{p \leq k} z_p + \lambda \sigma^2 \right) d\nu(\sigma). \quad (7.15)$$

Since X_0 is convex in λ , $\partial X_0 / \partial \lambda$ is increasing in λ and, therefore, (7.14) implies that $\lambda = \lambda(u)$ is nondecreasing in u . Therefore, in order to prove (7.13), it is enough to show that for some $\delta > 0$ if $\lambda(u) = 0$ then $u \in [d + \delta, D - \delta]$. We will only prove that $\lambda(u) = 0$ implies that $u \geq d + \delta$ as the case $u \leq D - \delta$ is quite similar.

We set $\lambda = 0$ and denote $Z = \sum_{p \leq k} z_p$. Given $\gamma > 0$, we can write

$$\mathbb{E} W_1 \dots W_k \langle \sigma^2 \rangle \geq \mathbb{E} W_1 \dots W_k \langle \sigma^2 \rangle I(|Z| \leq \gamma). \quad (7.16)$$

First of all, let us show that if $|Z| \leq \gamma$ then

$$\langle \sigma^2 \rangle \geq d + \frac{1}{L} \exp(-L\gamma), \quad (7.17)$$

where a constant L does not depend on γ (but it depends on other parameters of the model such as measure ν). To show this, let us first note that if $|Z| \leq \gamma$ we have

$$\exp X_{k+1} = \int_{\Sigma} \exp(\sigma Z) d\nu(\sigma) \leq L \exp(L\gamma), \quad (7.18)$$

using (1.7). Next,

$$\begin{aligned} \langle \sigma^2 \rangle \exp X_{k+1} &= \int_{\Sigma} \sigma^2 \exp(\sigma Z) d\nu(\sigma) \\ &= d \exp X_{k+1} + \int_{\Sigma} (\sigma^2 - d) \exp(\sigma Z) d\nu(\sigma) \\ &\geq d \exp X_{k+1} + \frac{1}{L} \exp(-L\gamma), \end{aligned} \quad (7.19)$$

where the last inequality is obtained by restricting the integral to the set $\{\sigma^2 \in [(D+d)/2, D]\}$ of positive measure ν by (1.7). Combining (7.18) and (7.19) yields (7.17). Plugging (7.17) into (7.16) implies

$$\begin{aligned} \mathbb{E} W_1 \dots W_k \langle \sigma^2 \rangle &\geq \left(d + \frac{1}{L} \exp(-L\gamma) \right) \mathbb{E} W_1 \dots W_k I(|Z| \leq \gamma) \\ &= \left(d + \frac{1}{L} \exp(-L\gamma) \right) \left(1 - \mathbb{E} W_1 \dots W_k I(|Z| > \gamma) \right). \end{aligned} \quad (7.20)$$

Using Hölder's inequality, we can bound

$$\begin{aligned}\mathbb{E}W_1 \dots W_k I(|Z| > \gamma) &\leq (\mathbb{P}(|Z| > \gamma))^{1/2} (\mathbb{E}(W_1 \dots W_k)^2)^{1/2} \\ &\leq L \exp\left(-\frac{\gamma^2}{L}\right) (\mathbb{E}(W_1 \dots W_k)^2)^{1/2},\end{aligned}\tag{7.21}$$

since Z is a Gaussian r.v. with variance $\xi(u)$ uniformly bounded for all $d \leq u \leq D$. As in (4.11) we can write

$$W_1 \dots W_k = \exp X_{k+1} \prod_{l \leq k} (\exp(-X_l))^{m_l - m_{l-1}}$$

and by Hölder's inequality,

$$\mathbb{E}(W_1 \dots W_k)^2 \leq (\mathbb{E} \exp(4X_{k+1}))^{1/2} \prod_{l \leq k} (\mathbb{E} \exp(-4X_l))^{(m_l - m_{l-1})/2}.\tag{7.22}$$

The first factor on the right hand side can be estimated as follows,

$$\begin{aligned}\mathbb{E} \exp(4X_{k+1}) &= \mathbb{E} \left(\int_{\Sigma} \exp(\sigma Z) d\nu(\sigma) \right)^4 \leq \mathbb{E} \int_{\Sigma} \exp(4\sigma Z) d\nu(\sigma) \\ &= \int_{\Sigma} \exp(8\sigma^2 \xi'(u)) d\nu(\sigma) \leq L.\end{aligned}$$

Next, since

$$X_l = \frac{1}{m_l} \log \mathbb{E}_l \exp m_l X_{l+1} \geq \mathbb{E}_l X_{l+1},$$

repeating this over l yields that $X_l \geq \mathbb{E}_l X_{k+1}$. Therefore,

$$\begin{aligned}\mathbb{E} \exp(-4X_l) &\leq \mathbb{E} \exp(-4\mathbb{E}_l X_{k+1}) \leq \mathbb{E} \exp(-4X_{k+1}) = \mathbb{E} \left(\int_{\Sigma} \exp(\sigma Z) d\nu(\sigma) \right)^{-4} \\ &\leq \mathbb{E} \int_{\Sigma} \exp(-4\sigma Z) d\nu(\sigma) = \int_{\Sigma} \exp(8\sigma^2 \xi'(u)) d\nu(\sigma) \leq L.\end{aligned}$$

Plugging all these estimates into (7.22) gives

$$\mathbb{E}(W_1 \dots W_k)^2 \leq L$$

and (7.21) implies

$$\mathbb{E}W_1 \dots W_k I(|Z| > \gamma) \leq L \exp\left(-\frac{\gamma^2}{L}\right).$$

Finally, (7.20) implies that

$$\begin{aligned}u = \mathbb{E}W_1 \dots W_k \langle \sigma^2 \rangle &\geq \left(d + \frac{1}{L} \exp(-L\gamma)\right) \left(1 - L \exp\left(-\frac{\gamma^2}{L}\right)\right) \\ &\geq d + \frac{1}{L} \exp(-L\gamma) - L \exp\left(-\frac{\gamma^2}{L}\right).\end{aligned}$$

Taking γ large enough gives $u \geq d + \delta$ for some $\delta > 0$. As we have already mentioned above one can similarly show that $\lambda = 0$ implies that $u \leq D - \delta$ for some $\delta > 0$ and this finishes

the proof of Lemma.

□

We are now ready to prove Theorem 8.

Proof of Theorem 8. Let us start with equation (5.6) that states that for any λ, \mathbf{m} and \mathbf{q} such that $q_{k+1} = u$,

$$F_N(u, \varepsilon_N) \leq -\lambda u_N(\lambda) + X_0(\mathbf{m}, \mathbf{q}, \lambda) - \frac{1}{2} \sum_{1 \leq l \leq k} m_l(\theta(q_{l+1}) - \theta(q_l)) + \mathcal{R}, \quad (7.23)$$

where $u_N(\lambda)$ is defined in (5.3) and $|\mathcal{R}| \leq c_N + L\varepsilon_N$. Let us note that the definition (5.3) implies that

$$|u_N(\lambda) - u| \leq \varepsilon_N. \quad (7.24)$$

Let us take $\delta > 0$ as in Lemma 11. Lemma 10 implies that for all $d + \delta \leq u \leq D - \delta$ we have $|\lambda(u)| \leq \Lambda(\delta)$. Moreover, (7.23) and (7.24) imply that for any $d + \delta \leq u \leq D - \delta$ and $|\lambda| \leq \Lambda(\delta)$ we have

$$\begin{aligned} F_N(u, \varepsilon_N) &\leq -\lambda u + X_0(\mathbf{m}, \mathbf{q}, \lambda) - \frac{1}{2} \sum_{1 \leq l \leq k} m_l(\theta(q_{l+1}) - \theta(q_l)) + \Lambda(\delta)\varepsilon_N + \mathcal{R} \\ &= \mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u) + \mathcal{R}, \end{aligned} \quad (7.25)$$

where again $|\mathcal{R}| \leq c_N + L\varepsilon_N$. The fact that for $d + \delta \leq u \leq D - \delta$ we have $|\lambda(u)| \leq \Lambda(\delta)$ means in this case that for a fixed (\mathbf{m}, \mathbf{q}) minimizing $\mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u)$ over λ is equivalent to minimizing it over $|\lambda| \leq \Lambda(\delta)$ and, therefore, minimizing $\mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u)$ over $(\mathbf{m}, \mathbf{q}, \lambda)$ is also equivalent to minimizing it over \mathbf{m}, \mathbf{q} and $|\lambda| \leq \Lambda(\delta)$. Therefore, (7.25) implies that for $d + \delta \leq u \leq D - \delta$,

$$F_N(u, \varepsilon_N) \leq \mathcal{P}(\xi, u) + \mathcal{R} \leq \sup_{d \leq u \leq D} \mathcal{P}(\xi, u) + \mathcal{R}, \quad (7.26)$$

where $|\mathcal{R}| \leq c_N + L\varepsilon_N$.

Next, let u be such that $u + \varepsilon_N < d + \delta$. Let us consider arbitrary $k' \geq 1$ and arbitrary \mathbf{m}' and \mathbf{q}' such that $q'_{k'+1} = u_N(\lambda)$. By (7.24), there exist \mathbf{m} and \mathbf{q} , maybe, with different parameter k , such that $q_{k+1} = u$ and

$$\int |m(q) - m'(q)| dq \leq |u - u_N(\lambda)| \leq \varepsilon_N, \quad (7.27)$$

where functions $m(q)$ and $m'(q)$ are defined in (A.32) in terms of (\mathbf{m}, \mathbf{q}) and $(\mathbf{m}', \mathbf{q}')$ correspondingly. This can be achieved by simply assigning $m(q) = 1$ for q between u and $u_N(\lambda)$ and, otherwise, letting $m(q) = m'(q)$. Then Lemma 9 implies that

$$|X_0(\mathbf{m}, \mathbf{q}, \lambda) - X_0(\mathbf{m}', \mathbf{q}', \lambda)| \leq L\varepsilon_N.$$

Also, obviously, condition (7.27) implies that

$$\left| \frac{1}{2} \sum_{1 \leq l \leq k} m_l(\theta(q_{l+1}) - \theta(q_l)) - \frac{1}{2} \sum_{1 \leq l \leq k'} m'_l(\theta(q'_{l+1}) - \theta(q'_l)) \right| \leq L\varepsilon_N.$$

Therefore, (7.23) implies that for any k' , \mathbf{m}' and \mathbf{q}' such that $q'_{k'+1} = u_N(\lambda)$,

$$\begin{aligned} F_N(u, \varepsilon_N) &\leq -\lambda u_N(\lambda) + X_0(\mathbf{m}', \mathbf{q}', \lambda) - \frac{1}{2} \sum_{1 \leq l \leq k'} m'_l (\theta(q'_{l+1}) - \theta(q'_l)) + \mathcal{R} \\ &= \mathcal{P}_{k'}(\mathbf{m}', \mathbf{q}', \lambda, u_N(\lambda)) + \mathcal{R}, \end{aligned} \quad (7.28)$$

where again $|\mathcal{R}| \leq c_N + L\varepsilon_N$. Since now $u + \varepsilon_N < d + \delta < D$, for $\lambda < 0$ the definition (5.3) implies that $u_N(\lambda) = u + \varepsilon_N$ and, therefore, for $\lambda < 0$ we have

$$F_N(u, \varepsilon_N) \leq \mathcal{P}_{k'}(\mathbf{m}', \mathbf{q}', \lambda, u + \varepsilon_N) + \mathcal{R}. \quad (7.29)$$

Since $u + \varepsilon_N < d + \delta$, Lemma 11 implies that for any \mathbf{m}' and \mathbf{q}' , $\lambda(u) < 0$ which means that for fixed $(\mathbf{m}', \mathbf{q}')$ minimizing $\mathcal{P}_{k'}(\mathbf{m}', \mathbf{q}', \lambda, u + \varepsilon_N)$ over λ is equivalent to minimizing it over $\lambda < 0$ and, therefore, minimizing $\mathcal{P}_{k'}(\mathbf{m}', \mathbf{q}', \lambda, u + \varepsilon_N)$ over $(\mathbf{m}', \mathbf{q}', \lambda)$ is also equivalent to minimizing it over \mathbf{m}', \mathbf{q}' and $\lambda < 0$. Hence, (7.29) yields that if $u + \varepsilon_N < d + \delta$ then,

$$F_N(u, \varepsilon_N) \leq \mathcal{P}(\xi, u + \varepsilon_N) + \mathcal{R} \leq \sup_{d \leq u \leq D} \mathcal{P}(\xi, u) + \mathcal{R}. \quad (7.30)$$

Similarly, one can show that this holds for u such that $u - \varepsilon_N > D - \delta$ and, combining this with (7.26), we showed that for all u in the set

$$\{u + \varepsilon_N < d + \delta\} \cup \{d + \delta \leq u \leq D - \delta\} \cup \{u - \varepsilon_N > D - \varepsilon_N\} \quad (7.31)$$

we have

$$F_N(u, \varepsilon_N) \leq \sup_{d \leq u \leq D} \mathcal{P}(\xi, u) + \mathcal{R}.$$

It is obvious that for $\varepsilon_N = N^{-1}$ the set (7.31) contains ε_N -net of the interval $[d, D]$ of cardinality LN . This finishes the proof of Theorem 8. \square

Acknowledgment. I would like to thank David Sherrington for suggesting the topic of this research.

References

- [1] Ghatak, S.K., Sherrington, D. (1977) Crystal field effects in a general S Ising spin glass, *J. Phys. C: Solid State Phys.* **10**, 3149.
- [2] Guerra, F. (2003) Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.* **233**, no. 1, 1-12.
- [3] Guerra, F., Toninelli, F.L. (2002) The thermodynamic limit in mean field spin glass models, *Comm. Math. Phys.* **230**, 71-79.
- [4] Mottishaw, P.J., Sherrington, D. (1985) Stability of a crystal-field split spin glass. *J. Phys. C: Solid State Phys.* **18**, 5201-5213.

- [5] Ledoux, M., Talagrand, M. (1991) Probability in Banach spaces. Isoperimetry and Processes. Springer-Verlag.
- [6] Panchenko, D. (2005) A question about Parisi functional. Preprint.
- [7] Sherrington, D., Kirkpatrick, S. (1972) Solvable model of a spin glass. *Phys. Rev. Lett.* **35**, 1792-1796.
- [8] Talagrand, M. (2003) Spin Glasses: a Challenge for Mathematicians. Springer-Verlag.
- [9] Talagrand, M. (2003) On Guerra's broken replica-symmetry bound. *C. R. Math. Acad. Sci. Paris* **337**, no. 7, 477-480.
- [10] Talagrand, M. (2003) The generalized Parisi formula. *C. R. Math. Acad. Sci. Paris* **337**, no. 2, 111-114.
- [11] Talagrand, M. (2003) Parisi formula. To appear in *Ann. Math.*
- [12] Talagrand, M. (2003) On the meaning of Parisi's functional order parameter. *C. R. Math. Acad. Sci. Paris* **337**, no. 9, 625-628.
- [13] Talagrand, M. (2004) Free energy of the spherical mean field model. To appear in *Probab. Theory Related Fields*.
- [14] Talagrand, M. (2004) Parisi measures. To appear in *J. Funct. Analysis*.

A Proof of the apriori estimates.

We will prove the apriori estimates of Section 6 in several steps. In Section A.1 we obtain the analogue of Talagrand's interpolation for two copies of the system that is the main technical tool of the proof. In Section A.2, we summarize several properties of the parameters in the definition of the Parisi formula in (1.15) and (1.16) and in Sections A.3 and A.4 we prove the apriori estimates of Section 6 by considering two separate cases of "far" points and "close" points.

A.1 Talagrand's interpolation for two copies.

The key to proving the apriori estimate of Section 6 is Talagrand's interpolation for two copies of the system. The proof of this result is similar to the proof of Guerra's interpolation in Theorem 4. Once the Talagrand's interpolation is obtained, the arguments in the rest of the paper will be adapted from [11] with some necessary modifications.

For $v \in [-D, D]$, let $v = \eta|v|$ for $\eta = \pm 1$. Consider $\kappa \geq 1$ and consider a sequence \mathbf{n} :

$$0 = \mathbf{n}_0 \leq \mathbf{n}_1 \leq \dots \leq \mathbf{n}_\kappa = 1 \tag{A.1}$$

such that $\mathbf{n}_\tau = |v|$ for some $\tau \leq \kappa$. Consider a sequence $\boldsymbol{\rho}$:

$$0 = \rho_0 \leq \dots \leq \rho_{\kappa+1} = u. \tag{A.2}$$

Consider a sequence of pairs of random variables (Y_p^1, Y_p^2) independent for $0 \leq p \leq \kappa$ such that

$$\mathbb{E}(Y_p^1)^2 = \mathbb{E}(Y_p^2)^2 = t(\xi'(\rho_{p+1}) - \xi'(\rho_p)) \quad (\text{A.3})$$

and such that

$$Y_p^1 = \eta Y_p^2 \text{ for } p < \tau \text{ and } Y_p^1, Y_p^2 \text{ are independent for } p \geq \tau. \quad (\text{A.4})$$

Let (Z_p^1, Z_p^2) be an arbitrary sequence of independent vectors for $0 \leq p \leq \kappa$. Let $(Z_{i,p}^1, Z_{i,p}^2)$ and $(Y_{i,p}^1, Y_{i,p}^2)$ be independent copies for $i \leq N$ of the sequences (Z_p^1, Z_p^2) and (Y_p^1, Y_p^2) and we assume that they are independent of each other and the randomness in the Hamiltonian $H_N(\boldsymbol{\sigma})$. For $s \in [0, 1]$, let us define an interpolating Hamiltonian

$$H_s(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) = \sqrt{st}H_N(\boldsymbol{\sigma}^1) + \sqrt{st}H_N(\boldsymbol{\sigma}^2) + \sum_{j \leq 2} \sum_{i \leq N} \sigma_i^j \sum_{p \leq \kappa} \left(Z_{i,p}^j + \sqrt{1-s} Y_{i,p}^j \right) \quad (\text{A.5})$$

and consider

$$F = \log \int_{A(v)} \exp H_s(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) d\nu(\boldsymbol{\sigma}^1) d\nu(\boldsymbol{\sigma}^2), \quad (\text{A.6})$$

where $A(v)$ was defined in (6.14). Let

$$\chi(s) = \frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F. \quad (\text{A.7})$$

Theorem 9 *For $s \in [0, 1]$ we have*

$$\chi'(s) \leq -2t \sum_{l < \tau} \mathbf{n}_l (\theta(\rho_{l+1}) - \theta(\rho_l)) - t \sum_{l \geq \tau} \mathbf{n}_l (\theta(\rho_{l+1}) - \theta(\rho_l)) + \mathcal{R} \quad (\text{A.8})$$

where $|\mathcal{R}| \leq K(\varepsilon_N + c(N))$.

Proof. The argument is similar to Guerra's interpolation in Theorem 4. (3.15) implies that

$$\chi'(s) = \frac{1}{N} \mathbb{E} W_1 \dots W_\kappa \frac{\partial F}{\partial s},$$

where

$$W_l = \exp \mathbf{n}_l (F_{l+1} - F_l)$$

and where F_l are defined as in (3.2). Using the fact that $\mathbf{n}_\kappa = 1$, one can write

$$\begin{aligned} W_1 \dots W_\kappa &= \exp \sum_{1 \leq l \leq \kappa} \mathbf{n}_l (F_{l+1} - F_l) \\ &= \exp \left(F + \sum_{1 \leq l \leq \kappa} (\mathbf{n}_{l-1} - \mathbf{n}_l) F_l \right) = T \exp F \end{aligned} \quad (\text{A.9})$$

where $T = T_1 \dots T_\kappa$ and $T_l = \exp(\mathbf{n}_{l-1} - \mathbf{n}_l)F_l$. Using the definition of F in (A.6),

$$\chi'(s) = \frac{1}{N} \mathbb{E} T \exp F \frac{\partial F}{\partial s} = \text{I} - \text{II}$$

where

$$\text{I} = \frac{\sqrt{t}}{2N\sqrt{s}} \int_{A(v)} \mathbb{E} T (H_N(\boldsymbol{\sigma}^1) + H_N(\boldsymbol{\sigma}^2)) \exp H_s(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) d\nu(\boldsymbol{\sigma}^1) d\nu(\boldsymbol{\sigma}^2). \quad (\text{A.10})$$

and

$$\text{II} = \frac{1}{2N\sqrt{1-s}} \int_{A(v)} \mathbb{E} T \sum_{j \leq 2} \sum_{i \leq N} \sigma_i^j \sum_{p \leq \kappa} Y_{i,p}^j \exp H_s(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) d\nu(\boldsymbol{\sigma}^1) d\nu(\boldsymbol{\sigma}^2). \quad (\text{A.11})$$

To simplify the notations, let us denote $\boldsymbol{\sigma} = (\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2)$ and $\boldsymbol{\rho} = (\boldsymbol{\rho}^1, \boldsymbol{\rho}^2)$, let $d\nu(\boldsymbol{\sigma}) = d\nu(\boldsymbol{\sigma}^1) d\nu(\boldsymbol{\sigma}^2)$ and define

$$\begin{aligned} \mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho}) &= \frac{1}{N} \mathbb{E} (H_N(\boldsymbol{\sigma}^1) + H_N(\boldsymbol{\sigma}^2)) (H_N(\boldsymbol{\rho}^1) + H_N(\boldsymbol{\rho}^2)) \\ &= \zeta(\boldsymbol{\sigma}^1, \boldsymbol{\rho}^1) + \zeta(\boldsymbol{\sigma}^1, \boldsymbol{\rho}^2) + \zeta(\boldsymbol{\sigma}^2, \boldsymbol{\rho}^1) + \zeta(\boldsymbol{\sigma}^2, \boldsymbol{\rho}^2) \end{aligned} \quad (\text{A.12})$$

where ζ was defined in (4.14). To compute I, we will use (4.9) for the family

$$\mathbf{g}(\boldsymbol{\rho}) = H_N(\boldsymbol{\rho}^1) + H_N(\boldsymbol{\rho}^2) \text{ for } \boldsymbol{\rho} = (\boldsymbol{\rho}^1, \boldsymbol{\rho}^2) \in A(v)$$

and we will think of each factor T_l in T as the functional of $\mathbf{g}(\boldsymbol{\rho})$. Then (4.9) and (A.12) imply,

$$\begin{aligned} \text{I} &= \frac{\sqrt{t}}{2\sqrt{s}} \int_{A(v)} \mathbb{E} T \frac{\partial \exp H_s(\boldsymbol{\sigma})}{\partial \mathbf{g}(\boldsymbol{\sigma})} \mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}) \\ &+ \frac{\sqrt{t}}{2\sqrt{s}} \sum_{l \leq \kappa} \int_{A(v)} \mathbb{E} T \exp H_s(\boldsymbol{\sigma}) \frac{1}{T_l} \frac{\delta T_l}{\delta \mathbf{g}} [\mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho})] d\nu(\boldsymbol{\sigma}). \end{aligned} \quad (\text{A.13})$$

First of all,

$$\frac{\partial \exp H_s(\boldsymbol{\sigma})}{\partial \mathbf{g}(\boldsymbol{\sigma})} = \sqrt{st} \exp H_s(\boldsymbol{\sigma}),$$

and, therefore, the first line in (A.13) can be written as

$$\frac{t}{2} \int_{A(v)} \mathbb{E} T \exp H_s(\boldsymbol{\sigma}) \mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}) = \frac{t}{2} \mathbb{E} W_1 \dots W_\kappa \langle \mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \rangle, \quad (\text{A.14})$$

where

$$\langle \mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \rangle = \exp(-F) \int_{A(v)} \mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \exp H_s(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}).$$

Using the definition of T_l we can write

$$\frac{1}{T_l} \frac{\delta T_l}{\delta \mathbf{g}} [\mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho})] = (\mathbf{n}_{l-1} - \mathbf{n}_l) \frac{\delta F_l}{\delta \mathbf{g}} [\mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho})] = (\mathbf{n}_{l-1} - \mathbf{n}_l) \mathbb{E}_l W_l \dots W_\kappa \frac{\delta F}{\delta \mathbf{g}} [\mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho})],$$

where we used (3.16). We have

$$\frac{\delta F}{\delta \mathbf{g}}[\mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho})] = \sqrt{st} \exp(-F) \int_{A(v)} \mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho}) \exp H_s(\boldsymbol{\rho}) d\nu(\boldsymbol{\rho}) =: \sqrt{st} \langle \mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho}) \rangle',$$

where $\langle \cdot \rangle'$ denotes the Gibbs average with respect to $\boldsymbol{\rho}$ for a fixed $\boldsymbol{\sigma}$. Therefore, for a fixed $\boldsymbol{\sigma}$ we get

$$\frac{1}{T_l} \frac{\delta T_l}{\delta \mathbf{g}}[\mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho})] = \sqrt{st} (\mathbf{n}_{l-1} - \mathbf{n}_l) \mathbb{E} W_1 \dots W_\kappa \langle \mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho}) \rangle' =: \sqrt{st} (\mathbf{n}_{l-1} - \mathbf{n}_l) \gamma_l(\mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho})),$$

where γ_l here is defined by analogy with (4.5). Combining this with the fact that

$$T \exp H_s(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}) = W_1 \dots W_\kappa \exp(-F) \exp H_s(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma})$$

we can write the second line in (A.13) as

$$\begin{aligned} & \frac{t}{2} \sum_{l \leq \kappa} (\mathbf{n}_{l-1} - \mathbf{n}_l) \mathbb{E} W_1 \dots W_\kappa \exp(-F) \int_{A(v)} \gamma_l(\mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho})) d\nu(\boldsymbol{\sigma}) \\ &= \frac{1}{2} \sum_{l \leq \kappa} (\mathbf{n}_{l-1} - \mathbf{n}_l) \mathbb{E} W_1 \dots W_{l-1} \gamma_l^{\otimes 2}(\mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho})) = \frac{t}{2} \sum_{l \leq \kappa} (\mathbf{n}_{l-1} - \mathbf{n}_l) \mu_l(\mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho})), \end{aligned} \quad (\text{A.15})$$

where μ_l here is defined analogously to (4.6). Combining (A.14) and (A.15) we get

$$\text{I} = \frac{t}{2} \mathbb{E} W_1 \dots W_\kappa \langle \mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \rangle + \frac{t}{2} \sum_{l \leq \kappa} (\mathbf{n}_{l-1} - \mathbf{n}_l) \mu_l(\mathfrak{z}(\boldsymbol{\sigma}, \boldsymbol{\rho})). \quad (\text{A.16})$$

Let us denote by $R_{j,j'}$ the overlap of $\boldsymbol{\sigma}^j$ and $\boldsymbol{\sigma}^{j'}$ and by $R^{j,j'}$ the overlap of $\boldsymbol{\sigma}^j$ and $\boldsymbol{\rho}^{j'}$. From (1.1) and (A.12),

$$\text{I} = \frac{t}{2} \mathbb{E} W_1 \dots W_\kappa \left\langle \sum_{j,j' \leq 2} \xi(R_{j,j'}) \right\rangle + \frac{t}{2} \sum_{l \leq \kappa} (\mathbf{n}_{l-1} - \mathbf{n}_l) \mu_l \left(\sum_{j,j' \leq 2} \xi(R^{j,j'}) \right) + \mathcal{R}, \quad (\text{A.17})$$

where $|\mathcal{R}| \leq 4c(N)$. Since the average $\langle \cdot \rangle$ is over the set $A(v)$, we have

$$|R_{j,j} - u| \leq \varepsilon_N \text{ and } |R_{1,2} - v| \leq \frac{1}{N}. \quad (\text{A.18})$$

Therefore,

$$\text{I} = t(\xi(u) + \xi(v)) + \frac{t}{2} \sum_{l \leq \kappa} (\mathbf{n}_{l-1} - \mathbf{n}_l) \mu_l \left(\sum_{j,j' \leq 2} \xi(R^{j,j'}) \right) + \mathcal{R}, \quad (\text{A.19})$$

where $|\mathcal{R}| \leq K(\varepsilon_N + c(N))$. The computation of II is very similar. For $p \leq \kappa$, let us define

$$\mathbf{g}_p(\boldsymbol{\sigma}) = \sum_{j \leq 2} \sum_{i \leq N} \sigma_i^j Y_{i,p}^j$$

so that $\Pi = \sum_{p \leq \kappa} \Pi(p)$ where

$$\Pi(p) = \frac{1}{2N\sqrt{1-s}} \int_{A(v)} \mathbb{E} T \mathbf{g}_p(\boldsymbol{\sigma}) \exp H_s(\boldsymbol{\sigma}) d\nu(\boldsymbol{\sigma}).$$

Let us define

$$\mathfrak{z}_p(\boldsymbol{\sigma}, \boldsymbol{\rho}) = \frac{1}{N} \mathbb{E} \mathbf{g}_p(\boldsymbol{\sigma}) \mathbf{g}_p(\boldsymbol{\rho}) = \sum_{j,j' \leq 2} R^{j,j'} \mathbb{E} Y_p^j Y_p^{j'}. \quad (\text{A.20})$$

Then one can repeat the computations leading to (A.16) with one important difference. One needs to note that F_l does not depend on the r.v. $Y_{i,p}$ for $l \leq p$ and, as a result, the summation in the second term will be over $p < l \leq \kappa$, i.e.

$$\Pi(p) = \frac{1}{2} \mathbb{E} W_1 \dots W_\kappa \langle \mathfrak{z}_p(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \rangle + \frac{1}{2} \sum_{p < l \leq \kappa} (\mathbf{n}_{l-1} - \mathbf{n}_l) \mu_l(\mathfrak{z}_p(\boldsymbol{\sigma}, \boldsymbol{\rho})) \quad (\text{A.21})$$

and

$$\Pi = \frac{1}{2} \mathbb{E} W_1 \dots W_\kappa \left\langle \sum_{p \leq \kappa} \mathfrak{z}_p(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \right\rangle + \frac{1}{2} \sum_{l \leq \kappa} (\mathbf{n}_{l-1} - \mathbf{n}_l) \mu_l \left(\sum_{p < l} \mathfrak{z}_p(\boldsymbol{\sigma}, \boldsymbol{\rho}) \right) \quad (\text{A.22})$$

Using (A.18) and (A.3), it is easy to see that the first term on the right hand side is

$$tu\xi'(u) + t|v|\xi'(|v|) + \mathcal{R}.$$

Using (A.3) and (A.4),

$$\sum_{p < l} \mathbb{E} (Y_p^j)^2 = t\xi'(q_l) \text{ and } \sum_{p < l} \mathbb{E} Y_p^1 Y_p^2 = t\eta\xi'(q_{l \wedge \tau}) = t\xi'(\eta q_{l \wedge \tau}),$$

where in the last equality we used the fact that ξ is even function. If we define

$$q_l^{j,j} = q_l \text{ and } q_l^{1,2} = \eta q_{l \wedge \tau} \quad (\text{A.23})$$

then $\sum_{p < l} \mathbb{E} Y_p^j Y_p^{j'} = t\xi'(q_l^{j,j'})$ and (A.22) can be written as

$$\Pi = tu\xi'(u) + t|v|\xi'(|v|) + \frac{1}{2} \sum_{l \leq \kappa} (\mathbf{n}_{l-1} - \mathbf{n}_l) \mu_l \left(\sum_{j,j' \leq 2} R^{j,j'} \xi'(q_l^{j,j'}) \right) + \mathcal{R}. \quad (\text{A.24})$$

Since $\theta(x) = x\xi'(x) - \xi(x)$ is also even, (A.19) and (A.24) imply that

$$\begin{aligned} \chi'(s) &= \text{I} - \Pi = -t\theta(u) - t\theta(v) + \frac{t}{2} \sum_{l \leq \kappa} (\mathbf{n}_l - \mathbf{n}_{l-1}) \sum_{j,j' \leq 2} \theta(q_l^{j,j'}) + \mathcal{R} \\ &\quad - \frac{1}{2} \sum_{l \leq \kappa} (\mathbf{n}_l - \mathbf{n}_{l-1}) \mu_l \left(\sum_{j,j' \leq 2} \left(\xi(R^{j,j'}) - R^{j,j'} \xi'(q_l^{j,j'}) + \theta(q_l^{j,j'}) \right) \right) \\ &\leq -t\theta(u) - t\theta(v) + \frac{t}{2} \sum_{l \leq \kappa} (\mathbf{n}_l - \mathbf{n}_{l-1}) \sum_{j,j' \leq 2} \theta(q_l^{j,j'}) + \mathcal{R} \\ &= -t\theta(u) - t\theta(v) + \sum_{j,j' \leq 2} \theta(q_{\kappa+1}^{j,j'}) - \frac{t}{2} \sum_{j,j' \leq 2} \sum_{l \leq \kappa} \mathbf{n}_l (\theta(q_{l+1}^{j,j'}) - \theta(q_l^{j,j'})) + \mathcal{R} \\ &= -\frac{t}{2} \sum_{j,j' \leq 2} \sum_{l \leq \kappa} \mathbf{n}_l (\theta(q_{l+1}^{j,j'}) - \theta(q_l^{j,j'})) + \mathcal{R}, \end{aligned}$$

since $q_{\kappa+1}^{j,j} = u$ and $q_{\kappa+1}^{1,2} = |v|$. To finish the proof it remains to use (A.23). We have

$$\sum_{j,j' \leq 2} (\theta(q_{l+1}^{j,j'}) - \theta(q_l^{j,j'})) = 2(\theta(q_{l+1}) - \theta(q_l)) \text{ for } l < \tau$$

and

$$\sum_{j,j' \leq 2} (\theta(q_{l+1}^{j,j'}) - \theta(q_l^{j,j'})) = \theta(q_{l+1}) - \theta(q_l) \text{ for } l \geq \tau,$$

since for $l \geq \tau$, $\theta(q_{l+1}^{1,2}) - \theta(q_l^{1,2}) = \theta(q_\tau) - \theta(q_\tau) = 0$.

□

We can bound $\chi(0)$ as follows. Using (A.18), for any $\lambda, \gamma \in \mathbb{R}$,

$$\begin{aligned} F|_{s=0} &\leq -2N\lambda u - N\gamma v + 2N|\lambda|\varepsilon_N + |\gamma| \\ &+ \log \int_{(\Sigma^N)^2} \exp \sum_{i \leq N} \left(\sum_{j \leq 2} \sigma_i^j \sum_{p \leq \kappa} (Z_{i,p}^j + Y_{i,p}^j) + \sum_{j \leq 2} \lambda(\sigma_i^j)^2 + \gamma \sigma_i^1 \sigma_i^2 \right) d\nu(\boldsymbol{\sigma}^1) d\nu(\boldsymbol{\sigma}^2) \\ &= -2N\lambda u - N\gamma v + 2N|\lambda|\varepsilon_N + |\gamma| + \sum_{i \leq N} F_i(\lambda, \gamma), \end{aligned}$$

where $F_i(\lambda, \gamma)$ are independent copies of

$$F(\lambda, \gamma) = \log \int_{\Sigma \times \Sigma} \exp \left(\sum_{j \leq 2} \sigma_j \sum_{p \leq \kappa} (Z_p^j + Y_p^j) + \sum_{j \leq 2} \lambda(\sigma_j)^2 + \gamma \sigma_1 \sigma_2 \right) d\nu(\sigma_1) d\nu(\sigma_2). \quad (\text{A.25})$$

(3.5), (3.6) and (3.7) now imply that

$$\chi(0) \leq -2\lambda u - \gamma v + 2|\lambda|\varepsilon_N + |\gamma|N^{-1} + \mathcal{P}(\mathbf{n})F(\lambda, \gamma) \quad (\text{A.26})$$

and we obtained the following.

Corollary 2 *If $\chi(s)$ and $F(\lambda, \gamma)$ are defined by (A.7) and (A.25) then for all $\lambda, \gamma \in \mathbb{R}$,*

$$\chi(1) \leq -2\lambda u - \gamma v + \mathcal{P}(\mathbf{n})F(\lambda, \gamma) - 2t \sum_{l < \tau} \mathbf{n}_l(\theta(\rho_{l+1}) - \theta(\rho_l)) - t \sum_{l \geq \tau} \mathbf{n}_l(\theta(\rho_{l+1}) - \theta(\rho_l)) + \mathcal{R} \quad (\text{A.27})$$

where $|\mathcal{R}| \leq K(\varepsilon_N + |\lambda|\varepsilon_N + |\gamma|N^{-1} + c(N))$.

Proof. The proof follows immediately by combining (A.8) and (A.26).

□

Remark. The remainder \mathcal{R} in Theorem 6 will be a result of application of (A.27). We will use it for $\lambda = \lambda(u)$ defined in (5.20) and, as in the proof of Lemma 6, $|\lambda(u)| \leq \Lambda$ for some Λ that depends only on ξ, ν and u . Below we will use (A.27) for $|\gamma| \leq L$ for some constant L independent of N, t and v . As a result, the remainder in Theorem 6, $|\mathcal{R}| \leq a_N$ for $a_N = K(\varepsilon_N + c(N) + N^{-1})$.

A.2 Properties of ε -minimizer.

In this section we will describe several properties of the sequence $(k, \mathbf{m}, \mathbf{q}, \lambda)$ in (6.13) that follow from its definition. Section 4 in [11] describes these properties in a very general setting, with no reference to the classical SK model and all the computations there apply to our case. The only difference is that in [11] certain generic computations were applied to the function $\log \chi x$ and here we will apply them to the function

$$x \rightarrow \log \int_{\Sigma} \exp(\sigma x + \lambda \sigma^2) d\nu(\sigma). \quad (\text{A.28})$$

We will not reproduce some of the generic computations in [11] that directly apply to our case.

Perturbing the sequences \mathbf{m} and \mathbf{q} . Let

$$q_{r-1} \leq a \leq q_r, \quad b = \xi'(a) \text{ and } m_{r-1} \leq m \leq m_r$$

and define new sequences \mathbf{m}', \mathbf{q}' by inserting m and a into \mathbf{m} and \mathbf{q} . Let us consider a sequence $(z'_p)_{p \leq k+1}$ of independent random variables such that $\mathbb{E}(z'_p)^2 = \xi'(q'_{p+1}) - \xi'(q'_p)$ or, expressing this explicitly in terms of b ,

$$\begin{aligned} \mathbb{E}(z'_p)^2 &= \xi'(q_{p+1}) - \xi'(q_p) \text{ for } p < r-1, \\ \mathbb{E}(z'_{r-1})^2 &= b - \xi'(q_{r-1}), \quad \mathbb{E}(z'_r)^2 = \xi'(q_r) - b, \\ \mathbb{E}(z'_{p+1})^2 &= \xi'(q_{p+1}) - \xi'(q_p) \text{ for } r \leq p \leq k. \end{aligned}$$

Let

$$F = \log \int_{\Sigma} \exp\left(\sigma \sum_{p \leq k+1} z'_p + \lambda \sigma^2\right) d\nu(\sigma)$$

and consider functions

$$T(m, b) = \mathcal{P}(\mathbf{m}') F \quad (\text{A.29})$$

and

$$\Phi(m, a) = -\lambda u + T(m, \xi'(a)) - \frac{1}{2} \sum_{l \leq k} m_l (\theta(q_{l+1}) - \theta(q_l)) - \frac{1}{2} (m - m_{r-1}) (\theta(q_r) - \theta(a)). \quad (\text{A.30})$$

Comparing with the definition (1.15) it is clear that

$$T(m, \xi'(a)) = X_0(\mathbf{m}', \mathbf{q}', \lambda) \text{ and } \Phi(m, a) = \mathcal{P}_{k+1}(\mathbf{m}', \mathbf{q}', \lambda, u) \quad (\text{A.31})$$

and, thus, T and Φ describe the behavior of X_0, \mathcal{P}_k when we perturb the set of parameters by adding an extra point. It will be very convenient to note that the functionals X_0 and \mathcal{P}_k depend on the sequences \mathbf{m}, \mathbf{q} only through the function $m(q)$ defined by

$$m(q) = m_l \text{ for } q_l \leq q \leq q_{l+1}, \quad (\text{A.32})$$

which is called the *functional order parameter*. Therefore, inserting parameters m and a can be visualized as the perturbation of $m(q)$, as shown in Figure 3. From this point of view, it

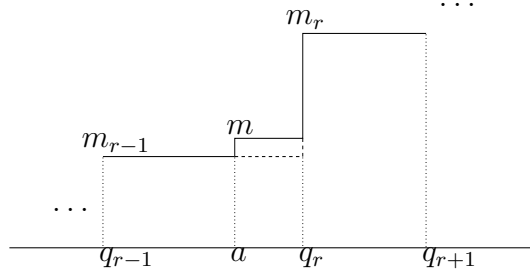


Figure 3: Perturbing the ε -minimizer.

becomes obvious that for all m and a as above,

$$T(m_{r-1}, b) = T(m, \xi'(q_r)) = X_0(\mathbf{m}, \mathbf{q}, \lambda), \quad (\text{A.33})$$

$$\Phi(m_{r-1}, a) = \Phi(m, q_r) = \mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u) \quad (\text{A.34})$$

and, thus, it is very important to study the behavior of the derivatives of T, Φ in m and a at $m = m_{r-1}$ and $a = q_r$.

Properties of the derivatives. Let us define

$$U(b) = 2 \frac{\partial T}{\partial m}(m, b) \Big|_{m=m_{r-1}} \quad (\text{A.35})$$

and

$$f(a) = \frac{\partial \Phi}{\partial m}(m, a) \Big|_{m=m_{r-1}} = \frac{1}{2} U(\xi'(a)) - \frac{1}{2} (\theta(q_r) - \theta(a)). \quad (\text{A.36})$$

The first fundamental formula is

$$\frac{\partial T}{\partial b}(m, b) \Big|_{b=\xi'(q_r)} = -\frac{1}{2} (m - m_{r-1}) A \quad (\text{A.37})$$

where A is independent of m . This can be obtained by a straightforward computation and one can write down an explicit formula for A (see [11]) but we will omit it here. By definition of the ε -minimizer (6.13), $(\mathbf{m}, \mathbf{q}, \lambda)$ is the minimizer of $\mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u)$ and, therefore,

$$\frac{\partial \mathcal{P}_k}{\partial q_r} = 0 \implies \frac{\partial X_0}{\partial q_r} + \frac{1}{2} (m_r - m_{r-1}) q_r \xi''(q_r) = 0, \quad (\text{A.38})$$

where we used (1.15) and the fact that $\theta'(q_r) = q_r \xi''(q_r)$. When $m = m_r$, it is apparent from Figure 3 that the intervals $[a, q_r]$ and $[q_r, q_{r+1}]$ are glued together in a sense that all functionals defined above become independent of q_r and, in particular, using (A.31)

$$T(m_r, \xi'(a)) = X_0(\mathbf{m}', \mathbf{q}', \lambda) \Big|_{m=m_r} = X_0(\mathbf{m}, \mathbf{q}, \lambda) \Big|_{q_r=a}. \quad (\text{A.39})$$

Taking the derivative of both sides with respect to a at $a = q_r$ gives

$$\frac{\partial X_0}{\partial q_r}(\mathbf{m}, \mathbf{q}, \lambda) = \frac{\partial T}{\partial a}(m_r, \xi'(a)) \Big|_{a=q_r} = \xi''(q_r) \frac{\partial T}{\partial b}(m_r, b) \Big|_{b=\xi'(q_r)} = -\frac{1}{2} (m_r - m_{r-1}) \xi''(q_r) A$$

where the last equality follows from (A.37) for $m = m_r$. Comparing this with (A.38) implies the first consequence of (6.13),

$$A = q_r. \quad (\text{A.40})$$

Next, (A.35) and (A.37) imply that

$$\frac{1}{2}U'(\xi'(q_r)) = \frac{\partial}{\partial b} \left(\frac{\partial T}{\partial m}(m, b) \Big|_{m=m_{r-1}} \right) \Big|_{b=\xi'(q_r)} = \frac{\partial}{\partial m} \left(\frac{\partial T}{\partial b}(m, b) \Big|_{b=\xi'(q_r)} \right) \Big|_{m=m_{r-1}} = -\frac{1}{2}A,$$

where, given any doubt, equality in the middle can be checked by computing both sides. Therefore,

$$-U'(\xi'(q_r)) = A = q_r. \quad (\text{A.41})$$

Another crucial property of $U(b)$ that can be verified by straightforward computation is

$$U''(b) \leq 0, \quad (\text{A.42})$$

i.e. $U(b)$ is concave in b . Next, let us describe several properties of $f(a)$ in (A.36). We have

$$f(q_r) = f'(q_r) = 0, \quad f(q_{r-1}) \geq 0. \quad (\text{A.43})$$

The first one follows from

$$f(q_r) = \frac{1}{2}U(\xi'(q_r)) = \frac{\partial T}{\partial m}(m, \xi'(q_r)) \Big|_{m=m_{r-1}} = 0,$$

since (A.33) yields that $T(m, \xi'(q_r))$ does not depend on m . The second one follows from (A.36) and (A.41) since

$$f'(q_r) = \frac{1}{2}\xi''(q_r)(U'(\xi'(q_r)) + q_r) = 0.$$

To show that $f(q_{r-1}) \geq 0$ let us note that

$$\Phi(m, q_{r-1}) = \mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u) \Big|_{m_{r-1}=m}$$

since setting $a = q_{r-1}$ simply replaces m_{r-1} with m in the definition of \mathcal{P}_k , which is also apparent from Figure 3. If $m_{r-1} > 0$, then as in (A.38),

$$\frac{\partial \mathcal{P}_k}{\partial m_{r-1}} = 0 \implies f(q_{r-1}) = \frac{\partial \Phi}{\partial m}(m, q_{r-1}) \Big|_{m=m_{r-1}} = 0.$$

If $m_{r-1} = 0$ then since $(\mathbf{m}, \mathbf{q}, \lambda)$ is the minimizer of \mathcal{P}_k , slightly increasing m_{r-1} should not decrease \mathcal{P}_k and, therefore, the right derivative $\partial \mathcal{P}_k / \partial m_{r-1} \geq 0$ and $f(q_{r-1}) \geq 0$.

ε -dependent properties of the derivatives. So far we have only utilized the fact that $(\mathbf{m}, \mathbf{q}, \lambda)$ is the minimizer of \mathcal{P}_k and we have not used the condition in (6.13) that

$$\mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u) \leq \mathcal{P}(\xi, u) + \varepsilon.$$

In particular, this implies that for any m and a ,

$$\Phi(m_{r-1}, a) = \mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u) \leq \mathcal{P}(\xi, u) + \varepsilon \leq \Phi(m, a) + \varepsilon, \quad (\text{A.44})$$

which means that we can not decrease \mathcal{P}_k much by varying parameters m, a . This can be combined with the following fact that plays a central role:

$$\begin{aligned} &\text{All derivatives of } T, \Phi, U, f \text{ with respect to } a, b, m \text{ are bounded by constants} \\ &\text{that depend only on } \xi, \nu, u \text{ but not on } (k, \mathbf{m}, \mathbf{q}). \end{aligned} \quad (\text{A.45})$$

Let L denote such constants that depend only on ξ, ν, u . The proof of (A.45) in [11] relied on the fact that $\mathbb{E} \text{ch}(c + c'z)^L \leq L$ for a standard Gaussian z and $|c|, |c'| \leq L$. In our case, this conditions will be replaced by an obvious condition,

$$\mathbb{E} \left(\int_{\Sigma} \exp(\sigma cz + \lambda \sigma^2) d\nu(\sigma) \right)^L \leq L \text{ for } |c|, |\lambda| \leq L.$$

The following Lemma holds.

Lemma 12 *The function $f(a)$ in (A.36) satisfies,*

$$f(a) \geq -L\sqrt{\varepsilon} \quad (\text{A.46})$$

and

$$f''(q_r) = \frac{1}{2} \xi''(q_r) (\xi''(q_r) U''(\xi'(q_r)) + 1) \geq -L\varepsilon^{1/6}. \quad (\text{A.47})$$

Proof. (A.46) holds if $f(a) \geq 0$ so we can that $f(a) < 0$. Using (A.36) and (A.45), we can write

$$\Phi(m, a) \leq \Phi(m_{r-1}, a) + (m - m_{r-1})f(a) + L(m - m_{r-1})^2. \quad (\text{A.48})$$

By (A.34) and the fact that $(\mathbf{m}, \mathbf{q}, \lambda)$ is a minimizer of \mathcal{P}_k ,

$$\Phi(m_{r-1}, a) = \mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u) \leq \mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, u)|_{q_r=a} = \Phi(m_r, a),$$

and the last equality can be seen as in (A.39). Therefore, (A.48) with $m = m_r$ implies that

$$(m_r - m_{r-1})f(a) + L(m_r - m_{r-1})^2 \geq 0$$

and, therefore,

$$m_r \geq m_{r-1} - \frac{f(a)}{L} \geq m_{r-1}, \quad (\text{A.49})$$

where the second inequality follows from our assumption that $f(a) < 0$. (A.44) and (A.48) imply that for any m and a ,

$$-\varepsilon \leq (m - m_{r-1})f(a) + L(m - m_{r-1})^2.$$

Taking $m := m_{r-1} - f(a)/2L$, which belongs to $[m_{r-1}, m_r]$ by (A.49), implies

$$-\varepsilon \leq -f(a)^2/4L$$

and this proves (A.46). It remains to prove (A.47). From (A.36) and (A.41) we see that

$$\begin{aligned} f''(q_r) &= \frac{1}{2} \xi'''(q_r) (U'(\xi'(q_r)) + q_r) + \frac{1}{2} \xi''(q_r) (U''(\xi'(q_r)) \xi''(q_r) + 1) \\ &= \frac{1}{2} \xi''(q_r) (U''(\xi'(q_r)) \xi''(q_r) + 1). \end{aligned}$$

(A.45) and (A.46) imply that for any a ,

$$-L\sqrt{\varepsilon} \leq f(a) \leq \frac{1}{2}f''(q_r)(a - q_r)^2 + L|a - q_r|^3. \quad (\text{A.50})$$

Using (A.43) we have

$$0 \leq f(q_{r-1}) \leq \frac{1}{2}f''(q_r)(q_{r-1} - q_r)^2 + L|q_{r-1} - q_r|^3$$

and, hence,

$$q_r \geq q_{r-1} - \frac{f''(q_r)}{2L}.$$

If $f''(q_r) \geq 0$ then (A.47) holds, otherwise

$$a = q_{r-1} - \frac{f''(q_r)}{2L} \in [q_{r-1}, q_r]$$

and using (A.50) for this choice of a again implies (A.47). □

Dual construction and the replica symmetric case.

The construction above will be used in the proof of Theorem 6 to provide control of the points on the left hand side of q_r , i.e. $q_{r-1} \leq v \leq q_r$. In order to provide control of the points on the right hand side $q_r \leq v \leq q_{r+1}$ one can consider a dual construction by perturbing parameter m_r on the interval $[q_r, a]$. This construction is very similar so we will not detail it and we will only consider the points on the left hand side in Theorem 6. In the replica symmetric case, the function $\Phi(m, a)$ defined in (2.7) is the analogue of the function in (A.30) and the properties (2.10) and (2.11) replace the properties (A.46) and (A.47) and, in fact, are stronger because ε is replaced by 0. (The change of sign in the inequalities is simply because we consider a dual construction.) Therefore, the proof of the replica symmetric apriori estimate in Theorem 7 is exactly the same as the proof of Theorem 6 if we use (2.10) and (2.11) instead of (A.46) and (A.47), and we will not detail it.

A.3 Control of the far points.

In this section we will prove the easiest case of Theorem 6, when the point v is far from q_r in the following sense. Given $\varepsilon > 0$, let $(k, \mathbf{m}, \mathbf{q}, \lambda)$ be an ε -minimizer defined by (6.13). Without loss of generality, we will assume that all coordinates of the vector \mathbf{m} are different and that all coordinates of the vector \mathbf{q} are also different. Otherwise, we can decrease the value of k by gluing equal coordinates without changing the value of the functional $\mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda)$. In this section we will consider the case when $v \in [-D, D]$ is such that

$$\begin{aligned} v < q_{r-1} \quad \text{or} \quad v > q_{r+1} \quad \text{if} \quad r \geq 2 \\ v < -q_1 \quad \text{or} \quad v > q_2 \quad \text{if} \quad r = 1 \end{aligned} \quad (\text{A.51})$$

and we will prove the following.

Proposition 2 *In the notations of Theorem 6, if (A.51) holds then*

$$\frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F(A(v)) \leq 2\psi(t) - K + \mathcal{R}, \quad (\text{A.52})$$

where $K > 0$ is a constant independent of N, t and v .

This implies Theorem 6 in the range of parameters (A.51) because

$$K \geq (v - q_r)^2 / (K^{-1}(q_r - q_{r-1})^2 \wedge (q_{r+1} - q_r)^2).$$

The proof of Proposition 2 is based on the following three step construction.

1. Let us recall the definition of \mathbf{n} in (6.4) and $(z_p^1, z_p^2)_{p \leq k}$ in (6.1).
2. (*Inserting $|v|$*). Let a be such that

$$q_a \leq |v| \leq q_{a+1}. \quad (\text{A.53})$$

Consider a vector

$$\mathbf{q}' = (q'_0, \dots, q'_{k+2}) = (q_0, \dots, q_{a-1}, |v|, q_a, \dots, q_{k+1})$$

which is defined by inserting $|v|$ in the vector \mathbf{q} and define a vector

$$\mathbf{n}' = (\frac{m_0}{2}, \dots, \frac{m_{a-1}}{2}, m_{a-1}, m_a, \dots, m_k).$$

Consider a sequence $(y_p)_{p \leq k+1}$ of independent Gaussian random variables such that

$$\mathbb{E} y_p^2 = \xi'(q'_{p+1}) - \xi'(q'_p)$$

and let $(y_p^1, y_p^2)_{p \leq k+1}$ consist of two copies of $(y_p)_{p \leq k+1}$ such that

$$y_p^1 = \eta y_p^2 \text{ for } p < a \text{ and } y_p^1, y_p^2 \text{ are independent for } p \geq a, \quad (\text{A.54})$$

where $v = \eta|v|$.

3. (*Gluing two sequences together*). Let $\kappa = 2k + 1$. Let us consider a vector

$$\mathbf{n} = (\mathbf{n}_0, \dots, \mathbf{n}_{\kappa+1}) \text{ such that } \mathbf{n}_0 \leq \dots \leq \mathbf{n}_{\kappa+1}$$

and such that \mathbf{n} consists of the elements of vectors \mathbf{n} and \mathbf{n}' . More precisely, there exists a partition I, J of the set $\{0, \dots, \kappa + 1\}$ with $\text{card} I = k + 1$ and $\text{card} J = k + 2$ such that the elements \mathbf{n}_p are the elements of \mathbf{n} for $p \in I$ and the elements of \mathbf{n}' for $p \in J$. For $0 \leq p \leq \kappa$, define

$$(Z_p^1, Z_p^2) = 0 \text{ for } p \in J \text{ and } (Z_p^1, Z_p^2) = \sqrt{1-t}(z_l^1, z_l^2) \text{ for } p \in I \quad (\text{A.55})$$

and where l is such that $\mathbf{n}_p = n_l$. Similarly, define

$$(Y_p^1, Y_p^2) = 0 \text{ for } p \in I \text{ and } (Y_p^1, Y_p^2) = \sqrt{t}(y_l^1, y_l^2) \text{ for } p \in J \quad (\text{A.56})$$

and where l is such that $\mathbf{n}_p = n'_l$. For $0 \leq p \leq \kappa$, let us define

$$(g_p^1, g_p^2) = (Z_p^1, Z_p^2) + (Y_p^1, Y_p^2) \quad (\text{A.57})$$

or, in other words,

$$(g_p^1, g_p^2) = (Z_p^1, Z_p^2) \text{ for } p \in I \text{ and } (g_p^1, g_p^2) = (Y_p^1, Y_p^2) \text{ for } p \in J.$$

In order to match this definition of (Y_p^j) with (A.2) and (A.3), let us define a sequence

$$\rho_0 \leq \dots \leq \rho_{\kappa+1}$$

such that $\rho_p = q'_l$ for l such that $\mathbf{n}_p = n'_l$. Define τ by $\mathbf{n}_\tau = n'_a = m_{a-1}$.

We will now apply Corollary 2 to these choices of \mathbf{n} , (Z_p^j) and (Y_p^j) . First of all,

$$\chi(1) = \frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F(A(v)) \quad (\text{A.58})$$

since for $s = 1$, the random variables (Y_p^j) will disappear in the definition of $H_s(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2)$ in (A.5), the Hamiltonian $H_s(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2)$ will coincide with Hamiltonian $H_t(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2)$ in (6.2) and, as a result, the definition of F in (A.6) will coincide with $F(A(v))$ in Theorem 6. Next, it is clear from the construction that

$$\begin{aligned} & -2t \sum_{l < \tau} \mathbf{n}_l (\theta(\rho_{l+1}) - \theta(\rho_l)) - t \sum_{l \geq \tau} \mathbf{n}_l (\theta(\rho_{l+1}) - \theta(\rho_l)) \\ &= -2t \sum_{l < a} n'_l (\theta(q'_{l+1}) - \theta(q'_l)) - t \sum_{l \geq a} n'_l (\theta(q'_{l+1}) - \theta(q'_l)) \\ &= -t \sum_{l \leq k} m_l (\theta(q_{l+1}) - \theta(q_l)). \end{aligned}$$

Corollary 2 now implies that

$$\frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F(A(v)) \leq -2\lambda u - \gamma v + \mathcal{P}(\mathbf{n}) F(\lambda, \gamma) - t \sum_{l \leq k} m_l (\theta(q_{l+1}) - \theta(q_l)) + \mathcal{R}.$$

We will use this bound for λ as in the ε -minimizer $(k, \mathbf{m}, \mathbf{q}, \lambda)$ and $\gamma = 0$, i.e.

$$\frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F(A(v)) \leq -2\lambda u + \mathcal{P}(\mathbf{n}) F(\lambda, 0) - t \sum_{l \leq k} m_l (\theta(q_{l+1}) - \theta(q_l)) + \mathcal{R}.$$

The argument in Lemma 6 shows that $|\lambda| \leq \Lambda$ for a constant Λ that depends only on ξ, ν and u and, hence, $|\mathcal{R}| \leq K(\varepsilon_N + c(N))$. Recalling the definition of $\psi(t)$ in (6.12), in order to prove Proposition 2, it remains to show that

$$\mathcal{P}(\mathbf{n}) F(\lambda, 0) \leq 2X_0(\mathbf{m}, \mathbf{q}, \lambda) - K = 2\mathcal{P}(\mathbf{m}) X_{k+1} - K,$$

where we used (3.4) and where $K > 0$ is a constant independent of t and v . In fact, it is enough to show that

$$\mathcal{P}(\mathbf{n}) F(\lambda, 0) < 2X_0(\mathbf{m}, \mathbf{q}, \lambda) = 2\mathcal{P}(\mathbf{m}) X_{k+1}, \quad (\text{A.59})$$

for all parameters $t \in [0, 1 - t_0]$ and v as in (A.51) because the functionals on both sides are continuous in these parameters and, even though the set defined in (A.51) is not a compact, the case of the end points will be proved in the following sections and (A.59) holds on the closure of (A.51). Therefore, by continuity and compactness, strict inequality for each (t, v) will imply strict inequality uniformly over the entire set of parameters. The proof of (A.59) repeats the proof of Proposition 5.7 in [11] with only one modification that instead of $\log \chi x$ we consider (A.28) and note that this function is also strictly convex in x because we eliminated the case when ν is concentrated on one point in Section B. Instead of reproducing the proof in its entirety we will explain a very clear idea behind it by looking at a few cases. Let us first show that a nonstrict version of (A.59), i.e.

$$\mathcal{P}(\mathbf{n})F(\lambda, 0) \leq 2X_0(\mathbf{m}, \mathbf{q}, \lambda) \quad (\text{A.60})$$

always holds, even without the assumption (A.51). (A.25) gives that $F(\lambda, 0) = F^1 + F^2$ where

$$F^j = \log \int_{\Sigma} \exp \left(\sigma \sum_{p \leq \kappa} g_p^j + \lambda \sigma^2 \right) d\nu(\sigma).$$

It is clear from the construction that for $p \in I$

$$\mathbf{n}_p = n_l = \frac{m_l}{2} \text{ for } l < r \iff g_p^1 = g_p^2 \quad (\text{A.61})$$

and for $p \in J$

$$\mathbf{n}_p = n'_l = \frac{m_l}{2} \text{ for } l < a \iff g_p^1 = \eta g_p^2. \quad (\text{A.62})$$

In other words, \mathbf{n}_p is of the type $m_l/2$ whenever the corresponding random pair is fully correlated. Let us define a vector $\mathbf{m} = (\mathbf{m}_0, \dots, \mathbf{m}_{\kappa})$ by

$$\mathbf{m}_p = 2\mathbf{n}_p \text{ for } p \text{ as in (A.61) or (A.62) and } \mathbf{m}_p = \mathbf{n}_p \text{ otherwise.} \quad (\text{A.63})$$

A fact that plays a very important role below is that coordinates of \mathbf{m} are not necessarily arranged in an increasing order. Let us first prove the following.

Lemma 13 *We have*

$$\mathcal{P}(\mathbf{n})F(\lambda, 0) = \mathcal{P}(\mathbf{n})(F^1 + F^2) \leq \mathcal{P}(\mathbf{m})F^1 + \mathcal{P}(\mathbf{m})F^2. \quad (\text{A.64})$$

Proof. This follows by induction in (3.2). For p such that $\mathbf{m}_p = \mathbf{n}_p$ and g_p^1, g_p^2 are independent we have

$$\begin{aligned} (F^1 + F^2)_p &= \frac{1}{\mathbf{n}_p} \log \mathbb{E}_p \exp \mathbf{n}_p (F^1 + F^2)_{p+1} \leq \frac{1}{\mathbf{n}_p} \log \mathbb{E}_p \exp \mathbf{n}_p (F_{p+1}^1 + F_{p+1}^2) \\ &= \frac{1}{\mathbf{m}_p} \log \mathbb{E}_p \exp \mathbf{m}_p F_{p+1}^1 + \frac{1}{\mathbf{m}_p} \log \mathbb{E}_p \exp \mathbf{m}_p F_{p+1}^2 = F_p^1 + F_p^2. \end{aligned} \quad (\text{A.65})$$

For p such that $\mathbf{m}_p = 2\mathbf{n}_p$ and $g_p^1 = \pm g_p^2$ we have

$$\begin{aligned} (F^1 + F^2)_p &= \frac{1}{\mathbf{n}_p} \log \mathbb{E}_p \exp \mathbf{n}_p (F^1 + F^2)_{p+1} \leq \frac{2}{\mathbf{m}_p} \log \mathbb{E}_p \exp \frac{\mathbf{m}_p}{2} (F_{p+1}^1 + F_{p+1}^2) \\ &\leq \frac{1}{\mathbf{m}_p} \log \mathbb{E}_p \exp \mathbf{m}_p F_{p+1}^1 + \frac{1}{\mathbf{m}_p} \log \mathbb{E}_p \exp \mathbf{m}_p F_{p+1}^2 = F_p^1 + F_p^2, \end{aligned} \quad (\text{A.66})$$

where in the second line we used Hölder's inequality. For $p = 0$ this gives (A.64). \square

Lemma 14 *Let \mathbf{m}' be a nondecreasing permutation of the vector \mathbf{m} . Then,*

$$\mathcal{P}(\mathbf{m})F^j \leq \mathcal{P}(\mathbf{m}')F^j = \mathcal{P}(\mathbf{m})X_{k+1} = X_0(\mathbf{m}, \mathbf{q}, \lambda). \quad (\text{A.67})$$

The first inequality means that the Parisi functional will decrease if \mathbf{m} is not arranged in an increasing order. Lemma 14 follows from Lemma 5.12 in [11] which states the following. Given a function Q and numbers $a \geq 0$ and $m > 0$, let

$$T_{m,a}(Q)(x) = \frac{1}{m} \log \mathbb{E} \exp mQ(x + g\sqrt{a})$$

where g is standard Gaussian.

Lemma 15 ([11]) *If $a, a' \geq 0$ and $m \geq m'$ then for each x we have*

$$T_{m,a} \circ T_{m',a'}(Q)(x) \leq T_{m',a'} \circ T_{m,a}(Q)(x). \quad (\text{A.68})$$

If $a, a' > 0$ and $m > m'$ then we can have equality only if Q is constant.

Proof of Lemma 14. The first inequality is obvious by Lemma 15. Equality in (A.67) follows by construction. The elements of \mathbf{m}' are precisely the elements of \mathbf{m} . The random variables g_p^j for p such that $\mathbf{m}'_p = m_l$ are exactly

$$\sqrt{1 - tz_l^j} \text{ and } \sqrt{ty_l^j} \text{ for } l \neq a - 1$$

and

$$\sqrt{1 - tz_l^j} \text{ and } \sqrt{ty_l^j}, \sqrt{ty_{l+1}^j} \text{ for } l = a - 1.$$

Obviously,

$$T_{m,a} \circ T_{m,a'} = T_{m,a+a'}, \quad (\text{A.69})$$

which means that we can combine the random variables corresponding to the same value m_l and since it is easy to check that in both cases the sum of these random variables is equal in distribution to z_l defined in (1.12), (A.67) follows. \square

Combining Lemma 13 and 14, we proved (A.60). From the proof it is clear that there are only two places, (A.66) and (A.68), where the inequality could become strict. It turns out that condition (A.51) ensures that gluing two sequences together occurs in such a way that at least in one of these two steps the inequality will become strict. We will not present the detailed proof here and refer a reader to Proposition 5.7 in [11]. We will explain the main idea by looking at several typical cases.

Let us consider the case $r \geq 2$ in (A.51). The case $r = 1$ is quite similar with the exception that the interval $-q_1 \leq v \leq q_0 = 0$ was excluded in (A.51) because it requires a different approach and it will be postponed until the following sections.

For $r \geq 2$ and v as in (A.51) we have

$$(a) |v| \notin [q_{r-1}, q_{r+1}] \quad \text{or} \quad (b) |v| \in [q_{r-1}, q_{r+1}] \text{ and } v = -|v|, \text{ i.e. } \eta = -1.$$

Case (a). Let us assume for simplicity that $q_{r-2} \leq |v| < q_{r-1}$ since other cases are similar. This corresponds to the case $a = r-2$ in (A.53). Then we will split case (a) into two subcases:

$$q_{r-2} \leq |v| < q_{r-1} \text{ and } m_{r-2} < m_{r-1}/2; \quad (\text{A.70})$$

$$q_{r-2} \leq |v| < q_{r-1} \text{ and } m_{r-1}/2 \leq m_{r-2}. \quad (\text{A.71})$$

Let us now see what happens when we combine the sequences at step 3 above. First of all, at step 1 the sequences \mathbf{n} and (z_p^1, z_p^2) will have subsequences

$$\begin{array}{ccccccccc} \dots & m_{r-2}/2 & m_{r-1}/2 & m_r & m_{r+1} & \dots & & & \\ \dots & (z_{r-2}, z_{r-2}) & (z_{r-1}, z_{r-1}) & (z_r^1, z_r^2) & (z_{r+1}^1, z_{r+1}^2) & \dots & & & \end{array} \quad (\text{A.72})$$

At step 2 the sequences \mathbf{n}' and (y_p^1, y_p^2) will have subsequences

$$\begin{array}{ccccccccc} \dots & m_{r-2}/2 & m_{r-2} & m_{r-1} & m_r & \dots & & & \\ \dots & (y_{r-2}, \eta y_{r-2}) & (y_{r-1}^1, y_{r-1}^2) & (y_r^1, y_r^2) & (y_{r+1}^1, y_{r+1}^2) & \dots & & & \end{array} \quad (\text{A.73})$$

In both cases we write (z^1, z^2) or (y^1, y^2) whenever two coordinates are independent. When we glue these sequences together as step 3, the sequences \mathbf{n} and (g_p^1, g_p^2) will contain subsequences

$$\begin{array}{ccccccccc} \dots & m_{r-2}/2 & m_{r-2}/2 & m_{r-2} & m_{r-1}/2 & m_{r-1} & \dots & & \\ \dots & (Z_{r-2}, Z_{r-2}) & (Y_{r-2}, \eta Y_{r-2}) & (Y_{r-1}^1, Y_{r-1}^2) & (Z_{r-1}, Z_{r-1}) & (Y_r^1, Y_r^2) & \dots & & \end{array} \quad (\text{A.74})$$

in the case (A.70) and

$$\begin{array}{ccccccccc} \dots & m_{r-2}/2 & m_{r-2}/2 & m_{r-1}/2 & m_{r-2} & m_{r-1} & \dots & & \\ \dots & (Z_{r-2}, Z_{r-2}) & (Y_{r-2}, \eta Y_{r-2}) & (Z_{r-1}, Z_{r-1}) & (Y_{r-1}^1, Y_{r-1}^2) & (Y_r^1, Y_r^2) & \dots & & \end{array} \quad (\text{A.75})$$

in the case (A.71).

Suppose that (A.74) is the case. Then the strict inequality will appear when we apply equation (A.66) at the step when \mathbf{n}_p is equal to $m_{r-1}/2$. Indeed, at this step

$$F_{p+1}^1 = F_{p+1}^1(\dots + Z_{r-2} + Y_{r-2} + Y_{r-1}^1 + Z_{r-1}),$$

$$F_{p+1}^2 = F_{p+1}^2(\dots + Z_{r-2} + Y_{r-2} + Y_{r-1}^2 + Z_{r-1})$$

and Y_{r-1}^1, Y_{r-1}^2 are independent and nondegenerate since $\mathbb{E}(Y_{r-1}^j)^2 = t(\xi'(q_{r-1}) - \xi'(|v|)) > 0$ by (A.70). Also, both functions $x \rightarrow F_{p+1}^j(x)$ are strictly convex because (A.28) is strictly convex and iteration in the Parisi functional (3.2) will preserve strict convexity. Therefore, F_{p+1}^1 and F_{p+1}^2 are not collinear as functions of Z_{r-1} with probability one over (Y_{r-1}^1, Y_{r-1}^2) and, therefore, Hölder's inequality in (A.66) will be strict with probability one.

Now, suppose that (A.75) holds. Then after using Lemma 13, $\mathcal{P}(\mathbf{m})F^1$ will be defined in terms of the sequences that contain subsequences

$$\begin{array}{ccccccc} \dots & m_{r-2} & m_{r-2} & m_{r-1} & m_{r-2} & m_{r-1} & \dots \\ \dots & Z_{r-2} & Y_{r-2} & Z_{r-1} & Y_{r-1}^1 & Y_r^1 & \dots \end{array} \quad (\text{A.76})$$

In this case, \mathbf{m} is not arranged in an increasing order, since $m_{r-1} > m_{r-2}$, and Z_{r-1}, Y_{r-1}^1 are nondegenerate. Therefore, when we rearrange these sequences in an increasing order by applying Lemma 15, we will get strict inequality in (A.68).

Case (b). In this case the scenario of (A.75) can not occur and the fact that $\eta = -1$ plays an important role. Suppose for certainty that $q_{r-1} \leq |v| \leq q_r$. (A.72) does not change but instead of (A.73) we will now have:

$$\begin{array}{ccccccc} \dots & m_{r-2}/2 & m_{r-1}/2 & m_{r-1} & m_r & \dots \\ \dots & (y_{r-2}, -y_{r-2}) & (y_{r-1}, -y_{r-1}) & (y_r^1, y_r^2) & (y_{r+1}^1, y_{r+1}^2) & \dots \end{array} \quad (\text{A.77})$$

When we glue this sequence with (A.72) we will get

$$\begin{array}{ccccccc} \dots & m_{r-2}/2 & m_{r-2}/2 & m_{r-1}/2 & m_{r-1}/2 & m_{r-1} & \dots \\ \dots & (Z_{r-2}, Z_{r-2}) & (Y_{r-2}, -Y_{r-2}) & (Z_{r-1}, Z_{r-1}) & (Y_{r-1}, -Y_{r-1}) & (Y_r^1, Y_r^2) & \dots \end{array} \quad (\text{A.78})$$

The strict inequality will appear when we apply equation (A.66) at the step when \mathbf{n}_p is equal to $m_{r-1}/2$ and

$$\begin{aligned} F_{p+1}^1 &= F_{p+1}^1(\dots + Z_{r-2} + Y_{r-2} + Z_{r-1} + Y_{r-1}), \\ F_{p+1}^2 &= F_{p+1}^2(\dots + Z_{r-2} - Y_{r-2} + Z_{r-1} - Y_{r-1}). \end{aligned}$$

Random variables Y_{r-2}, Z_{r-2} are independent and nondegenerate and we can argue as in the case (A.74) above. All other cases in the proof of Proposition 5.6 in [11] are very similar and (A.52) holds.

A.4 Control of the close points.

In Section A.3 we obtained the control of the points v far from q_r and in this section we will consider the remaining cases when $q_{r-1} \leq v \leq q_{r+1}$ or $-q_1 \leq v < 0$ when $r = 1$. All arguments repeat the arguments of Section 5 in [11], so we will only consider the case when

$$q_{r-1} \leq v \leq q_r.$$

As in the previous section, let L_1, L_2, \dots denote constants that depend only on ν, ξ and u . Consider a function

$$\Gamma(c) = \inf \left\{ |\xi(y) - \xi(x) + (x - y)\xi'(y)| : 0 \leq x, y \leq D, |x - y| \geq c \right\}. \quad (\text{A.79})$$

Since $\xi''(x) > 0$ we have $\Gamma(c) > 0$ for $c > 0$. In the notations of Theorem 6 the following holds.

Proposition 3 Suppose that $q_{r-1} \leq v \leq q_r$. If $L_1 \varepsilon^{1/6} \leq 1 - t_0$ then

$$L_1(q_r - v) \leq 1 - t_0 \Rightarrow \frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F(A(v)) \leq 2\psi(t) - \frac{(1 - t_0)^2}{L_1} (v - q_r)^2 + \mathcal{R}, \quad (\text{A.80})$$

and if $L_2 \varepsilon^{1/2} \leq (1 - t_0) \Gamma((1 - t_0)/L_1)$ then

$$L_1(q_r - v) \geq 1 - t_0 \Rightarrow \frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F(A(v)) < 2\psi(t) + \mathcal{R}. \quad (\text{A.81})$$

Together with a similar result for $q_r \leq v \leq q_{r+1}$ and the results of Section A.3, this proves Theorem 6. We will again use Talagrand's interpolation for two copies. Given

$$\frac{m_{r-1}}{2} \leq m \leq m_r,$$

let us define sequences \mathbf{n} and $\boldsymbol{\rho}$ in (A.1) and (A.2) by

$$0 = \mathbf{n}_0 = \frac{m_0}{2}, \mathbf{n}_1 = \frac{m_1}{2}, \dots, \mathbf{n}_{r-1} = \frac{m_{r-1}}{2}, \mathbf{n}_r = m, \mathbf{n}_{r+1} = m_r, \dots, \mathbf{n}_{k+1} = m_k$$

and

$$\rho_0 = q_0, \dots, \rho_{r-1} = q_{r-1}, \rho_r = v, \rho_{r+1} = q_r, \dots, \rho_{k+2} = q_{k+1}.$$

Since $\rho_r = v$, we have $\tau = r$. Consider a sequence (Y_p^1, Y_p^2) as in (A.3), i.e.

$$\mathbb{E}(Y_p^j)^2 = t(\xi'(\rho_{p+1}) - \xi'(\rho_p)),$$

$$Y_p^1 = Y_p^2 \text{ for } p < r \text{ and } Y_p^1, Y_p^2 \text{ are independent for } p \geq r.$$

Let (Z_p^1, Z_p^2) be such that

$$\mathbb{E}(Z_p^j)^2 = (1 - t)(\xi'(\rho_{p+1}) - \xi'(\rho_p))$$

for $p < r - 1$ and $p > r$,

$$\mathbb{E}(Z_{r-1}^j)^2 = (1 - t)(\xi'(q_r) - \xi'(q_{r-1}))$$

and $Z_r^j = 0$. Let

$$Z_p^1 = Z_p^2 \text{ for } p < r \text{ and } Z_p^1, Z_p^2 \text{ are independent for } p > r.$$

If we denote

$$g_p^j = Y_p^j + Z_p^j \text{ for } 0 \leq p \leq k + 1, j = 1, 2 \quad (\text{A.82})$$

then it follows from the construction that

$$\mathbb{E}(g_p^j)^2 = \xi'(\rho_{p+1}) - \xi'(\rho_p) \text{ for } p < r - 1, p > r,$$

$$\mathbb{E}(g_{r-1}^j)^2 = (\xi'(q_r) - \xi'(q_{r-1})) - t(\xi'(q_r) - \xi'(v)), \quad \mathbb{E}(g_r^j)^2 = t(\xi'(q_r) - \xi'(v)) \quad (\text{A.83})$$

and

$$g_p^1 = g_p^2 \text{ for } p < r \text{ and } g_p^1, g_p^2 \text{ - independent for } p > r.$$

If we define a point $a \in [v, q_r]$ by

$$\xi'(a) = t\xi'(v) + (1-t)\xi'(q_r) \quad (\text{A.84})$$

then (A.83) can be rewritten as

$$\mathbb{E}(g_{r-1}^j)^2 = \xi'(a) - \xi'(q_{r-1}), \quad \mathbb{E}(g_{r-1}^j)^2 = \xi'(q_r) - \xi'(a).$$

If we define a new sequence $\boldsymbol{\rho}'$ by

$$\rho'_0 = q_0, \dots, \rho'_{r-1} = q_{r-1}, \rho'_r = a, \rho'_{r+1} = q_r, \dots, \rho'_{k+2} = q_{k+1}$$

obtained by inserting a into the sequence \mathbf{q} , we finally get

$$\mathbb{E}(g_p^j)^2 = \xi'(\rho'_{p+1}) - \xi'(\rho'_p) \text{ for all } p \leq k+1 \quad (\text{A.85})$$

and

$$g_p^1 = g_p^2 \text{ for } p < r \text{ and } g_p^1, g_p^2 \text{ are independent for } p > r. \quad (\text{A.86})$$

Plugging the definition (A.82) into (A.25) we get

$$F(\lambda, \gamma) = \log \int_{\Sigma \times \Sigma} \exp \left(\sum_{j \leq 2} \sigma_j \sum_{p \leq k+1} g_p^j + \sum_{j \leq 2} \lambda(\sigma_j)^2 + \gamma \sigma_1 \sigma_2 \right) d\nu(\sigma_1) d\nu(\sigma_2). \quad (\text{A.87})$$

Let us define

$$V(\gamma, m, v) = \mathcal{P}(\mathbf{n})F(\lambda, \gamma) \quad (\text{A.88})$$

where we made the dependence of the right hand side on the parameters (γ, m, v) explicit. In order to apply Corollary 2, let us first note that from the construction of sequences \mathbf{n} and $\boldsymbol{\rho}$ we have

$$\begin{aligned} & -2t \sum_{l < \tau} \mathbf{n}_l(\theta(\rho_{l+1}) - \theta(\rho_l)) - t \sum_{l \geq \tau} \mathbf{n}_l(\theta(\rho_{l+1}) - \theta(\rho_l)) \\ &= -t \sum_{l \leq r-2} m_l(\theta(q_{l+1}) - \theta(q_l)) - t m_{r-1}(\theta(v) - \theta(q_{r-1})) \\ & \quad - t m(\theta(q_r) - \theta(v)) - t \sum_{r \leq l \leq k} m_l(\theta(q_{l+1}) - \theta(q_l)) \\ &= -t \sum_{l \leq k} m_l(\theta(q_{l+1}) - \theta(q_l)) - t(m - m_{r-1})(\theta(q_r) - \theta(v)). \end{aligned}$$

Corollary 2 and (A.58) now imply that

$$\begin{aligned} \frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F(A(v)) &\leq -2\lambda u - \gamma v + V(\gamma, m, v) \\ & \quad - t \sum_{l \leq k} m_l(\theta(q_{l+1}) - \theta(q_l)) - t(m - m_{r-1})(\theta(q_r) - \theta(v)) + \mathcal{R}. \end{aligned} \quad (\text{A.89})$$

One can easily check using the argument of Lemma 1 that

$$V(0, m, v) = 2T(m, \xi'(a)) \quad (\text{A.90})$$

where T was defined in (A.29). (A.33) implies that

$$V(0, m_{r-1}, v) = 2X_0(\mathbf{m}, \mathbf{q}, \lambda)$$

and, therefore, (A.89) with $\gamma = 0, m = m_{r-1}$ implies that

$$\frac{1}{N} \mathbb{E} \mathcal{P}(\mathbf{n}) F(A(v)) \leq 2\psi(t) + \mathcal{R}.$$

In order to prove Proposition 3, we will perturb parameters m and γ around these values $0, m_{r-1}$ and use the properties of ε -minimizer from the previous section. The fundamental connection of the bound (A.89) to the properties of ε -minimizer lies in the following fact:

$$\left. \frac{\partial V}{\partial \gamma}(\gamma, m_{r-1}, v) \right|_{\gamma=0} = -U'(\xi'(a)). \quad (\text{A.91})$$

The proof follows from straightforward computation and is given in Lemma 5.8 in [11]. Also, similar to (A.45), we have

$$\left| \frac{\partial^2 V}{\partial \gamma^2} \right| \leq L. \quad (\text{A.92})$$

We are now ready to prove Proposition 3.

Proof of Proposition 3. Let us consider a function

$$\alpha(\gamma) = V(\gamma, m_{r-1}, v) - \gamma v,$$

which is the part of the bound (A.89) for $m = m_{r-1}$ that depends on γ . By (A.91) we have

$$h(v) := \alpha'(0) = -U'(\xi'(a)) - v = -U'(t\xi'(v) + (1-t)\xi'(q_r)) - v,$$

since a was defined in (A.84). By (A.41) we have $h(q_r) = 0$ and

$$h'(q_r) = -t\xi''(q_r)U''(\xi'(q_r)) - 1. \quad (\text{A.93})$$

Using (A.47),

$$\begin{aligned} h'(q_r) &= -\xi''(q_r)U''(\xi'(q_r)) - 1 + (1-t)\xi''(q_r)U''(\xi(q_r)) \\ &\leq \frac{L}{\xi''(q_r)}\varepsilon^{1/6} + (1-t)\xi''(q_r)U''(\xi(q_r)). \end{aligned} \quad (\text{A.94})$$

We will now show that

$$L\varepsilon^{1/6} \leq 1 - t_0 \implies h'(q_r) \leq -\frac{1 - t_0}{4}. \quad (\text{A.95})$$

If $-\xi''(q_r)U''(\xi'(q_r)) \leq 1/2$ then (A.93) gives

$$h'(q_r) \leq \frac{t}{2} - 1 \leq -\frac{1}{2}.$$

If $-\xi''(q_r)U'''(\xi'(q_r)) \geq 1/2$ then

$$\frac{1}{\xi''(q_r)} \leq -2U'''(\xi'(q_r)) \leq L$$

by (A.45), and (A.94) gives

$$h'(q_r) \leq L\varepsilon^{1/6} - \frac{1-t}{2} \leq -\frac{1-t_0}{4},$$

where the last inequality holds if $4L\varepsilon^{1/6} \leq 1-t_0$. This proves (A.95). Since (A.45) implies that $|h''(v)| \leq L$ and since $h(q_r) = 0$, we can write

$$h(v) \geq (v - q_r)h'(q_r) - L(v - q_r)^2 \geq \frac{1}{8}(q_r - v)(1 - t_0)$$

if $q_r - v \leq (1 - t_0)/8L$ and if (A.95) holds. (A.92) implies that $|\alpha''(\gamma)| \leq L$ and we finally get

$$\begin{aligned} \inf_{\gamma} \alpha(\gamma) &\leq \inf_{\gamma} (\alpha(0) + \alpha'(0)\gamma + L\gamma^2) \leq \alpha(0) - \frac{\alpha'(0)^2}{L} = \alpha(0) - \frac{h(v)^2}{L} \\ &\leq 2X_0(\mathbf{m}, \mathbf{q}, \lambda) - \frac{1}{L}(1 - t_0)^2(v - q_r)^2. \end{aligned} \quad (\text{A.96})$$

Applying this to the bound (A.89) proves (A.80). Note that the infimum was achieved on $\gamma = -\alpha'(0)/2L$ and that (A.45) implies that $|\gamma| \leq L$. As we explained in the remark following Corollary 2, the bound (A.27) is used only for $|\gamma| \leq L$.

Next, we will prove (A.81). If $-h(v) = U'(\xi'(a)) + v \neq 0$ then we can simply use the first inequality in (A.96). Let us assume now that $U'(\xi'(a)) = -v$. Let us set $\gamma = 0$ in the bound (A.89) and consider the derivative of this bound in m at $m = m_{r-1}$, i.e.

$$\begin{aligned} D(t) &= \frac{\partial V}{\partial m}(0, m, v)|_{m=m_{r-1}} - t(\theta(q_r) - \theta(v)) \\ &= U(\xi'(a)) - t(\theta(q_r) - \theta(v)) && \text{by (A.90) and (A.35)} \\ &= U(t\xi'(v) + (1-t)\xi'(q_r)) - t(\theta(q_r) - \theta(v)). && \text{by (A.84)} \end{aligned}$$

Since we assumed that $U'(\xi'(a)) = -v$,

$$\begin{aligned} D'(t) &= (\xi'(v) - \xi'(q_r))U'(\xi'(a)) - (\theta(q_r) - \theta(v)) \\ &= -(\xi'(v) - \xi'(q_r))v - (\theta(q_r) - \theta(v)) \\ &= \xi(q_r) - \xi(v) + (v - q_r)\xi'(q_r) \leq -\Gamma(q_r - v), \end{aligned}$$

where the last inequality follows from the definition (A.79). By (A.42), $D(t)$ is concave in t and, therefore,

$$D(1) \leq D(t) + (1-t)D'(t) \leq D(t) - (1-t)\gamma(q_r - v).$$

By (A.46),

$$D(1) = U(\xi'(v)) - (\theta(q_r) - \theta(v)) = 2f(v) \geq -L\varepsilon^{1/2}.$$

We get

$$D(t) \geq (1-t)\gamma(q_r - v) - L\varepsilon^{1/2} > 0,$$

if $(1-t_0)\gamma(q_r - v) > L\varepsilon^{1/2}$, which is true under the conditions in (A.81) and one can finish the proof as in (A.96). \square

B Cases reducible to the classical SK model.

We will now show that only the case of $d < u < D$ in Theorem 1 is different from the classical SK model. First of all, $d = D$ means that $\Sigma = \{-\sqrt{d}, +\sqrt{d}\}$ which is precisely the case of the SK model. If measure ν has nonzero mass at both points $\pm\sqrt{d}$ then $\nu(\sigma)$ is proportional to $\exp h\sigma$ for some external field parameter $h \in \mathbb{R}$. Otherwise, if ν is concentrated at one point, the statement of Theorem 1 becomes trivial.

It remains to consider the cases of $d < D$ and $u = d$ or $u = D$. We will only consider the case $u = d$, since the case $u = D$ is similar. Let us consider a set

$$U_N(\varepsilon) = \{\sigma : R_{1,1} \in [d, d + \varepsilon]\}$$

and a function

$$F_N(\varepsilon) = \frac{1}{N} \mathbb{E} \log \int_{U_N(\varepsilon)} \exp H_N(\sigma) d\nu(\sigma).$$

Since, by (1.1),

$$\mathbb{E} \int_{\Sigma^N} \exp H_N(\sigma) d\nu(\sigma) < \infty,$$

the function $\exp H_N(\sigma)$ is ν -integrable with respect to σ almost surely and, therefore, by the monotone convergence theorem,

$$\lim_{\varepsilon \rightarrow 0} \int_{U_N(\varepsilon)} \exp H_N(\sigma) d\nu(\sigma) = \int_{\{R_{1,1}=d\}} \exp H_N(\sigma) d\nu(\sigma) \quad \text{a.s.}$$

Using the monotone convergence theorem once again implies

$$\lim_{\varepsilon \rightarrow 0} F_N(\varepsilon) = P_N := \frac{1}{N} \mathbb{E} \log \int_{\{R_{1,1}=d\}} \exp H_N(\sigma) d\nu(\sigma). \quad (\text{B.1})$$

If we choose a sequence (ε_N) so that $|F_N(\varepsilon_N) - P_N| \leq N^{-1}$, in order to prove Theorem 1 for $u = d$, it is enough to show that

$$\lim_{N \rightarrow \infty} P_N = \mathcal{P}(\xi, d). \quad (\text{B.2})$$

We will prove this by considering two separate cases.

Case 1. $\nu(\{\sigma^2 = d\}) = 0$. This means that the measure ν has no atoms at the points $\pm\sqrt{d}$ and, therefore, $\nu(\{R_{1,1} = d\}) = 0$ and $P_N = -\infty$. To prove (B.2), we need to show

that $\mathcal{P}(\xi, d) = -\infty$. Let us, for example, take $k = 1$ and $q_1 = 0$. For this choice of \mathbf{m}, \mathbf{q} we have

$$\begin{aligned} X_0(\mathbf{m}, \mathbf{q}, \lambda) &= \log \mathbb{E} \int_{\Sigma} \exp\left(\sigma z \sqrt{\xi'(u)} + \lambda \sigma^2\right) d\nu(\sigma). \\ &= \log \int_{\Sigma} \exp\left(\frac{1}{2} \sigma^2 \xi'(u) + \lambda \sigma^2\right) d\nu(\sigma). \end{aligned}$$

Therefore,

$$-\lambda d + X_0(\mathbf{m}, \mathbf{q}, \lambda) = \log \int_{\Sigma} \exp\left(\frac{1}{2} \sigma^2 \xi'(u) + \lambda(\sigma^2 - d)\right) d\nu(\sigma)$$

and, by the monotone convergence theorem,

$$\lim_{\lambda \rightarrow -\infty} (-\lambda d + X_0(\mathbf{m}, \mathbf{q}, \lambda)) = \log \int_{\{\sigma^2 = d\}} \exp\left(\frac{1}{2} \sigma^2 \xi'(u)\right) d\nu(\sigma) = -\infty,$$

since we assumed that $\nu(\{\sigma^2 = d\}) = 0$. Clearly, (1.16) yields that $\mathcal{P}(\xi, d) = -\infty$.

Case 2. $\nu(\{\sigma^2 = d\}) > 0$. This means that measure ν has at least one atom at the points $\pm\sqrt{d}$. Consider a probability measure $\bar{\nu}$ defined by

$$\bar{\nu}(\{-\sqrt{d}\}) = \frac{\nu(\{-\sqrt{d}\})}{\nu(\{\sigma^2 = d\})}, \quad \bar{\nu}(\{\sqrt{d}\}) = \frac{\nu(\{\sqrt{d}\})}{\nu(\{\sigma^2 = d\})}. \quad (\text{B.3})$$

Condition (1.7) implies that $R_{1,1} = d$ if and only if $\sigma_i^2 = d$ for all $i \leq N$. In other words,

$$\{R_{1,1} = d\} = \Sigma_d^N, \quad \text{where} \quad \Sigma_d = \{-\sqrt{d}, \sqrt{d}\}.$$

With these notations, P_N in (B.1) can be written as

$$P_N = \log \nu(\{\sigma^2 = d\}) + \bar{P}_N \quad (\text{B.4})$$

where

$$\bar{P}_N = \frac{1}{N} \mathbb{E} \log \int_{\Sigma_d^N} \exp H_N(\boldsymbol{\sigma}) d\bar{\nu}(\boldsymbol{\sigma}).$$

But \bar{P}_N exactly falls into the case $d = D$ considered above because $\bar{\nu}(\{\sigma^2 = d\}) = 1$, and, therefore, its limit can be written as follows. If we consider

$$\bar{Y}_{k+1} = \log \int_{\Sigma_d} \exp\left(\sigma \sum_{0 \leq p \leq k} z_p\right) d\bar{\nu}(\sigma),$$

define \bar{Y}_l recursively as in (1.14) and let

$$\bar{\mathcal{P}}_k(\mathbf{m}, \mathbf{q}) = \bar{Y}_0 - \frac{1}{2} \sum_{1 \leq l \leq k} m_l (\theta(q_{l+1}) - \theta(q_l))$$

then

$$\lim_{N \rightarrow \infty} \bar{P}_N = \inf \bar{\mathcal{P}}_k(\mathbf{m}, \mathbf{q}),$$

where the infimum is over all choices of k, \mathbf{m} and \mathbf{q} , which by (B.4) implies

$$\lim_{N \rightarrow \infty} P_N = \log \nu(\{\sigma^2 = d\}) + \inf \bar{\mathcal{P}}_k(\mathbf{m}, \mathbf{q}).$$

Equivalently, this can be written as follows. If we consider

$$Y_{k+1} = \log \nu(\{\sigma^2 = d\}) + \log \int_{\Sigma_d} \exp\left(\sigma \sum_{0 \leq p \leq k} z_p\right) d\bar{\nu}(\sigma), \quad (\text{B.5})$$

define Y_l recursively as in (1.14) and let

$$\mathcal{P}_k(\mathbf{m}, \mathbf{q}) = Y_0 - \frac{1}{2} \sum_{1 \leq l \leq k} m_l (\theta(q_{l+1}) - \theta(q_l)) \quad (\text{B.6})$$

then

$$\lim_{N \rightarrow \infty} P_N = \inf \mathcal{P}_k(\mathbf{m}, \mathbf{q}). \quad (\text{B.7})$$

By definition (B.3) of measure $\bar{\nu}$, Y_{k+1} in (B.5) can be also written as

$$Y_{k+1} = \log \int_{\{\sigma^2 = d\}} \exp\left(\sigma \sum_{0 \leq p \leq k} z_p\right) d\nu(\sigma). \quad (\text{B.8})$$

In order to prove (B.2), we will show that $\mathcal{P}(\xi, d)$ is equal to the right hand side of (B.7). The definition of $\mathcal{P}(\xi, d)$ given by (1.13) - (1.16) can be written equivalently as follows. If we consider

$$X_{k+1} = \log \int_{\Sigma} \exp\left(\sigma \sum_{0 \leq p \leq k} z_p + \lambda(\sigma^2 - d)\right) d\nu(\sigma), \quad (\text{B.9})$$

define X_l recursively as in (1.14), and define

$$\mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, d) = X_0 - \frac{1}{2} \sum_{1 \leq l \leq k} m_l (\theta(q_{l+1}) - \theta(q_l)), \quad (\text{B.10})$$

then

$$\mathcal{P}(\xi, d) = \inf \mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, d), \quad (\text{B.11})$$

where the infimum is taken over all λ, k, \mathbf{m} and \mathbf{q} . Since $d \leq \sigma^2$ for $\sigma \in \Sigma$, X_{k+1} in (B.9) is increasing in λ which implies that X_0 is also increasing in λ . Therefore, for any fixed \mathbf{m} and \mathbf{q} to minimize the right hand side of (B.11) over λ one should let $\lambda \rightarrow -\infty$. By the monotone convergence theorem, almost surely,

$$\lim_{\lambda \rightarrow -\infty} X_{k+1} = \int_{\{\sigma^2 = d\}} \exp\left(\sigma \sum_{0 \leq p \leq k} z_p\right) d\nu(\sigma) = Y_{k+1}.$$

Using the monotone convergence theorem repeatedly in the recursive construction (1.14) gives $\lim_{\lambda \rightarrow -\infty} X_0 = Y_0$ and comparing (B.6) and (B.10) we get

$$\inf_{\lambda} \mathcal{P}_k(\mathbf{m}, \mathbf{q}, \lambda, d) = \mathcal{P}_k(\mathbf{m}, \mathbf{q}).$$

Combining this with (B.7) and (B.11) gives (B.2). □